Karl-Heinz Loesgen

Das Bayessche Theorem und Modelle mit Fehlern in den Variablen

Dissertation zur Erlangung des Grades
eines Doktors der Wirtschaftswissenschaften
der Rechts- und Wirtschaftswissenschaftlichen
Fakultät der Universität des Saarlandes
(doctor rerum politicarum)

1974

Karl-Heinz Loesgen

Das Bayessche Theorem und Modelle mit Fehlern in den Variablen

Dissertation zur Erlangung des Grades
eines Doktors der Wirtschaftswissenschaften
der Rechts- und Wirtschaftswissenschaftlichen
Fakultät der Universität des Saarlandes
(doctor rerum politicarum)

UNIV.
LEIDEN
BIBL.

Tag der Disputation: 14. Februar 1974

Dekan: Prof. Dr. jur. Günter Jahr

Erstberichterstatter: Prof. Dr. phil. nat. Hans Schneeweiß

Zweitberichterstatter: Prof. Dr. rer. pol. Werner Neubauer

Das Bayessche Theorem und Modelle mit Fehlern in den Variablen

Inhalt:

		Seite
1	Einleitende Erörterungen	1
1.1	Ökonomische Relevanz des Problems der Fehler in den Variablen	1
1.2	A priori Information und Parameterschätzung	2
1.3	Das Bayessche Theorem	3
2	Vages a priori Wissen	6
2.1	Multiples Modell mit voneinander abhängenden Fehlern in den Variablen	6
2.2	Multiples Modell mit unabhängigen Fehlern in den Beobachtungen	12
2.3	Einfache Regression	16
3	Echtes a priori Wissen	20
3.1	A priori Information über die Regressionskoeffizienten	20
3.2	A priori Information über die Kovarianzmatrix der Fehler	22
3.3	A priori Information über die Regressionskoeffizienten und die Kovarianzmatrix der Fehler	32
3.4	A priori Information über die fehlerhaft beobachteten erklärenden Variablen	35
4	Übersicht über die gewonnenen Ergebnisse	42
Anhang		
1	Zur Integration über die erklärenden fehlerhaft beobachteten Variablen	47
2	Einige Integralprobleme und damit in Zusammenhang stehende Wahrscheinlichkeitsdichtefunktionen	53
Literaturverzeichnis		60

1 Einleitende Erörterungen

1.1 Ökonometrische Relevanz des Problems der Fehler in den Variablen

Die Ökonometrie ist in der Verfolgung ihres Zieles, wirtschaftliche Zusammenhänge empirisch zu erforschen - nicht nur qualitative sondern sogar quantitative Aussagen über die Dependenz bzw. Interdependenz ökonomischer Variablen zu machen - im ganz überwiegenden Teil ihrer Untersuchungen auf statistisches Zahlenmaterial angewiesen, über dessen Güte und Genauigkeit wenig oder nichts bekannt ist. So werden zum Beispiel in den im Zentrum des Interesses stehenden Analysen makroökonomischer Modelle Zahlen der Volkswirtschaftlichen Gesamtrechnung entnommen und meist wird - häufig genug kritiklos - angenommen, diese Daten seien exakt richtig, obwohl doch jedem Ökonometriker bekannt ist, daß mit Sicherheit zumindest einige der dort vorliegenden Zeitreihen mit Fehlern behaftet sind [1] oder aber gar nicht genau den Größen adäquat sind, die eigentlich aus wirtschaftstheoretischen Gründen in das Modell gehören sollten[2]. Des weiteren ist man hin und wieder auf Daten angewiesen, die lediglich aufgrund von Stichproben geschätzt worden sind.[3] In diesem Fall kann man aber immerhin aus dem

---

1) Untersuchungen über die Genauigkeit der Volkswirtschaftlichen Gesamtrechnung enthalten die Arbeiten von: S.Kuznets(1954),P.Studenski(1961), O.Morgenstern(1965) und S.S.Zarkovich(1966). Eine neuere Arbeit über Fehler in den Beobachtungen lieferte K.Weichselberger(1970).

2) In der vorliegenden Arbeit beschränke ich mich auf die Betrachtung zufälliger Fehler, die im Mittel Null sein sollen; eine Annahme, die sicherlich hinsichtlich vieler Makrovariablen problematisch ist.

3) Siehe z.B. M.H.Hansen, W.N.Hurwitz und M.A. Bershad (1961)

Stichprobenergebnis - sofern die entsprechenden Voraussetzungen der Stichprobentheorie erfüllt sind - außer den Schätzungen der Variablen auch Informationen über die Verteilung dieser Schätzungen oder anders gesehen über die Verteilung der Fehler in den Variablen erhalten. Insbesondere läßt sich hier die Kovarianzmatrix dieser Fehler schätzen.

1.2 A priori Information und Parameterschätzung

In der Tat sind über die Datenreihen und die Spezifikation des Modells mit seinen stochastischen Voraussetzungen hinausgehende Informationen notwendig, will man im Falle von Fehlern in den Beobachtungen zu Parameterschätzungen gelangen, über deren Eigenschaften noch positive Aussagen möglich sind.

Die Methode der kleinsten Quadrate führt unter bestimmten Regularitätsbedingungen hinsichtlich der Matrix der Realisationen der erklärenden Variablen zu inkonsistenten Schätzungen. Der Grund hierfür liegt darin, daß die Beobachtungen der erklärenden Variablen mit der Störvariablen der Regression der Beobachtungen der zu erklärenden Variablen auf die Beobachtungen der erklärenden Variablen korreliert sind und entsprechend der Wahrscheinlichkeitslimes des durch den Stichprobenumfang dividierten Skalarproduktes der Matrix der fehlerhaften Beobachtungen mit dem Vektor der Realisationen der Störvariablen von Null verschieden ist. In diesem Sinne sind die fehlerhaften Beobachtungen auch der exogenen Variablen nicht mehr als exogen zu bezeichnen.[4)]

Das Maximum Likelihood Prinzip versagt im Falle der Normalverteilungshypothese über die Beobachtungsfehler und den Fehler in der Gleichung sofern wir keine Information über die Kovarianzmatrix der Fehler haben insoweit als die notwendigen Bedingungen für ein Extremum

---

4) Vgl. H. Schneeweiß (1971) S. 220

Beziehungen zwischen den Schätzungen postulieren, die unter den "wahren" Parametern nicht bestehen.[5)]

Haben wir dagegen bestimmte a priori Informationen, so ist es möglich, unverzerrte Schätzungen zu gewinnen. Kennen wir z.B. die Kovarianzmatrix der Fehler, so führt im Falle der Normalverteilungsannahme das Maximum Likelihood Prinzip zu konsistenten Schätzungen.[6)] Andere - aber eben auch a priori - Voraussetzungen müssen erfüllt sein, damit das Waldsche Verfahren der Gruppierung der Beobachtungen [7)] oder die Methode der Instrumentvariablen[8)] sinnvoll angewendet werden können.

1.3 Das Bayessche Theorem

Nun ist in der Realität vollständige a priori Information - vollständig im Sinne der exakten Kenntnis eines Parameters oder eines theoretischen Momentes - wohl nur in Ausnahmefällen zu erhalten. Auch in dem angesprochenen Beispiel, in dem die Fehler in den Variablen reine Stichprobenfehler sind, ist die Varianz bzw. die Kovarianzmatrix im multiplen Fall eben nur zu schätzen, wobei allerdings bei großen Stichproben die Wahrscheinlichkeit einer Abweichung vom richtigen Wert klein sein wird.

Eine systematische Einbeziehung unvollständiger a priori Information, dergestalt daß über die unbekannten Parameter eine a priori Wahrscheinlichkeitsdichtefunktion gegeben ist, gestattet das Bayessche Theorem[9)], sofern aufgrund der Modellannahmen die Likelihoodfunktion berechnet werden kann. Das Bayessche Theorem folgt direkt

---

5) Siehe z.B. H. Johnston (1963) S. 152

6) ebenda S. 168-170 u. H. Schneeweiß (1971) S.224-227

7) A. Wald (1940)

8) Siehe z.B. H. Johnston (1963) S. 165 f.

9) T. Bayes (1763/64)

aus der Definition der bedingten Wahrscheinlichkeit: Für zwei Ereignisse A_k und B gelte:

$$(1.1) \quad P(A_k \mid B) = \frac{P(A_k B)}{P(B)}$$

sofern B nicht mit Wahrscheinlichkeit Null eintritt. Wendet man diese Definition umgekehrt auf B gegeben A_k an, löst nach $P(A_k B)$ auf und setzt das Ergebnis in (1.1) ein, so erhält man:

$$(1.2) \quad P(A_k \mid B) = \frac{P(B \mid A_k) P(A_k)}{P(B)} \quad .$$

Bezeichnen wir mit $A_1, A_2, \ldots, A_k, \ldots$ Ereignisse, für die gelten soll: $A_i A_j = \emptyset$ für $i \neq j$ sowie $\Sigma_{i=1}^{\infty} A_i = S$, wobei $\emptyset$ das unmögliche und S das sichere Ereignis sein soll, so können wir den Nenner von (1.2) weiter umformen:

$$(1.3) \quad P(B) = P(\Sigma_{i=1}^{\infty} B A_i) = \Sigma_{i=1}^{\infty} P(B A_i) = \Sigma_{i=1}^{\infty} P(B \mid A_i) P(A_i),$$

und wir erhalten den Satz von Bayes:

$$(1.4) \quad P(A_k \mid B) = \frac{P(B \mid A_k) P(A_k)}{\Sigma_{i=1}^{\infty} P(B \mid A_i) P(A_i)} \quad .$$

Diese Beziehung läßt sich auf den Fall kontinuierlicher Zufallsvariablen übertragen, für die die entsprechenden Wahrscheinlichkeitsdichtefunktionen existieren:[10)]

$$(1.5) \quad p_1(A \mid B) = c \; p(B \mid A) \; p_o(A) \quad , \text{ wobei}$$

$$(1.6) \quad c^{-1} = \int_{\mathbb{R}} p(B \mid A) \; p_o(A) \; dA \quad .$$

---

10) Siehe z.B. A. Rényi (1962) S. 231 f.

In dieser Form wird das Bayessche Theorem im folgenden benutzt. Dabei ist B als eine Matrix von Beobachtungen und a priori gegebenen Konstanten zu interpretieren, während unter A die Menge der unbekannten Parameter des Modells zu verstehen ist. $p_o(A)$, die a priori Dichtefunktion, enthält dabei die a priori Information, soweit sie unvollständig ist, also Parameter nicht exakt bekannt sind. $p(B|A)$ ist die Dichtefunktion der Beobachtungen bei gegebener Menge von Parametern. Betrachtet man die Beobachtungen als gegeben und die Parameter als variabel, so heißt $p(B|A)$ Likelihoodfunktion. c ist dann als eine Konstante zu verstehen, da c nur noch von den gegebenen Beobachtungen B, nicht mehr von A abhängt. c läßt sich stets wie in (1.6) berechnen, sofern die a posteriori Dichte $p_1(A|B)$ eine echte Wahrscheinlichkeitsdichtefunktion ist. Daher beschränke ich mich in den folgenden Untersuchungen auf die Betrachtung des Produktes aus Likelihood und a priori Dichtefunktion als Ausgangspunkt.

Es werden nacheinander verschiedene a priori Wahrscheinlichkeitsverteilungen angenommen, die - angefangen mit dem Fall völliger Informationslosigkeit über die Parameter - a priori Kenntnisse über die Koeffizienten, die Kovarianzmatrix der Fehler und die fehlerhaft beobachteten erklärenden Variablen unterstellen. Dabei wird nach der Berechnung der a posteriori Wahrscheinlichkeitsverteilung der Koeffizienten insbesondere jeweils die Rolle der nach der gewöhnlichen Methode der kleinsten Quadrate errechneten Schätzungen herausgearbeitet, insofern als Parameter, die die a posteriori Verteilung bestimmen, in K-Q-Schätzungen ausgedrückt werden. Weiter wird in den Fällen, in denen dies sinnvoll erscheint, das asymptotische Verhalten von Parametern der a posteriori Verteilung untersucht.

2 Vages a priori Wissen

Trotz der in Abschnitt 1.2 hervorgehobenen Notwendigkeit von a priori Kenntnissen in Modellen mit Fehlern in den Variablen möchte ich zunächst untersuchen, welche Resultate die Bayessche Methode im Falle fehlender a priori Information liefert.[11)]

2.1 Multiples Modell mit voneinander abhängenden Fehlern in den Variablen

Das multiple Modell mit Fehlern in den Variablen hat folgende Gestalt:

(2.1) $$X_2 = \Xi + V$$

(2.2) $$y = \eta + w$$

(2.3) $$\eta = X_1\alpha + \Xi\beta + \varepsilon$$

Dabei ist η der T-Spaltenvektor der Realisationen, y der T-Spaltenvektor der Beobachtungen und w der T-Spaltenvektor der Fehler in den Beobachtungen der zu erklärenden Variablen. Ξ ist die T × n Matrix der Realisationen, X_2 die T × n Matrix der Beobachtungen und V die T × n Matrix der Fehler in den Beobachtungen der fehlerhaft gemessenen erklärenden Variablen. X_1 ist die T × m Matrix exakt beobachteter erklärender Variablen und ε ist der T-Spaltenvektor der Fehler in der Gleichung. Schließlich sind α und β die m bzw. n-Spaltenvektoren der Regressionskoeffizienten.

Die Annahme exakt gemessener erklärender Variablen gestattet die Deutung von (2.3) als inhomogene Regression, wenn eben X_1 eine Scheinvariable enthält.

Definieren wir noch

(2.4) $$\omega := w + \varepsilon$$

so können wir folgende Annahmen über die stochastischen

---

11) Wie die a priori Wahrscheinlichkeitsdichtefunktion aussieht, die fehlender a priori Information entsprechen soll, wird von Fall zu Fall erläutert.

Eigenschaften des Modells formulieren:

(A1) (v_t', ω_t) sei normalverteilt mit dem Nullvektor als Mittelwert und der Kovarianzmatrix Σ für $t=1,2,\ldots,T$, wobei v_t' die t. Zeile von V und ω_t die t. Komponente von ω ist. Σ sei positiv definit.

(A2) (v_t', ω_t) sei stochastisch unabhängig von vorhergehenden Fehlern für $t = 1,2,\ldots,T$.

Die erste Annahme ist sicherlich am wenigsten kontrovers. Da wir zur Anwendung des Bayesschen Theorems die Likelihoodfunktion benötigen, ist die Festlegung des Typs der Verteilung der Fehlervariablen notwendig und die Annahme der Normalverteilung ist naheliegend, da wir in ökonometrischen Untersuchungen zumeist Makrovariablen betrachten und wir uns die Fehler in den Beobachtungen großer Aggregate durch Addition vieler kleiner unabhängiger Einzelfehler zustande gekommen denken können. Haben diese Einzelfehler endliche Varianz, so ist der Gesamtfehler zumindest annähernd normalverteilt.

Die zweite Annahme stellt eine stärkere Einschränkung der Anwendbarkeit des Modells dar, denn gerade in Relationen unter ökonomischen Variablen spricht manches für die Annahme von Autokorrelation in den Beobachtungsfehlern und den Fehlern in der Gleichung.

Aus den Annahmen folgt:

$$f(V,\omega|\Sigma) \propto |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2} sp(V,\omega)\ \Sigma^{-1}\ (V,\omega)'} . \tag{2.5}$$

Zur Formulierung der Likelihoodfunktion fassen wir unter Benutzung von (2.4) die Gleichungen (2.2) und (2.3) zusammen zu:

$$y = X_1\alpha + E\beta + \omega . \tag{2.6}$$

Und gemäß (2.1) und (2.6) wird (V,ω) in (X_2,y) transformiert, was keine Schwierigkeiten bereitet, da die Jakobische Determinante gleich 1 ist:

$$(2.7) \qquad f(X_2,y|\alpha,\beta,\Sigma,\Xi,X_1) \propto |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2}\,sp\,A}$$

wobei: $A := (X_2-\Xi, y-X_1\alpha-\Xi\beta)\Sigma^{-1}(X_2-\Xi, y-X_1\alpha-\Xi\beta)'$.

Zur Anwendung des Bayesschen Theorems benötigen wir noch die Spezifikation der a priori Wahrscheinlichkeitsdichte über die unbekannten Parameter α,β,Σ und Ξ. Ich möchte die a priori Dichte zunächst nur hinsichtlich Ξ festlegen: Mangelnder a priori Information über die Elemente von Ξ wird am besten die Annahme einer diffusen Gleichverteilung gerecht. Für eine solche "Verteilung" sprechen Invarianzargumente[12], die Forderung nach konjugierten Verteilungen[13] aber auch informationstheoretische Überlegungen: Eine solche diffuse Verteilung hat eben unter bestimmten recht allgemeinen Annahmen maximale Entropie bzw. sie enthält minimale Information im Sinne des Shannonschen Informationsmaßes[14]. Weiterhin läßt sich zeigen, daß diese diffuse Verteilung im Mittel den durch die Stichprobe möglichen Informationsgewinn maximiert[15].

Es sei also a priori folgende Annahme getroffen:

$$(2.8) \qquad p_o(\alpha,\beta,\Sigma,\Xi) \propto p_o(\alpha,\beta,\Sigma).$$

Nach dem Bayesschen Theorem ergibt sich aus (2.7) und (2.8) als a posteriori Dichte:

$$(2.9) \qquad p_1(\alpha,\beta,\Sigma,\Xi|y,X_1X_2) \propto p_o(\alpha,\beta,\Sigma)|\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2}\,sp\,A}.$$

Da wir vor allem Aussagen über die Regressionskoeffizienten α und β gewinnen wollen, ist das Ziel der weiteren Analyse die Berechnung der a posteriori Randverteilung von α und β. Die a posteriori Randdichtefunktion gewinnen

---

12) H. Jeffreys (1966) S.117 und J. Hartigan (1964)
13) H. Raiffa und R. Schlaifer (1961) S.58-69
14) C.E. Shannon (1948) oder A. Rényi (1962) S. 440
15) A. Zellner (1971) S. 41-53

wir durch Integration von (2.9) über die anderen Parameter[16]. Zunächst soll also das bestimmte Integral von (2.9) über die Elemente von Ξ über den Raum der reellen Zahlen gelöst werden. Zur Abkürzung der - wie wir sehen werden - etwas ausufernden Schreibarbeit sei definiert:

$$(2.10)\quad y^{\times} = y - X_1\alpha \quad .$$

Die Spur von A läßt sich in folgende Form bringen:

$$\text{sp } A = \text{sp}(\Xi - \hat{\Xi})(I_n \beta)\,\Sigma^{-1}(I_n \beta)'(\Xi-\hat{\Xi})' + \text{sp}(X_2, y^{\times})\{\Sigma^{-1} - \Sigma^{-1}(I_n \beta)'\left[(I_n \beta)\,\Sigma^{-1}(I_n \beta)\right]^{-1}(I_n \beta)\,\Sigma^{-1}\}(X_2, y^{\times})',$$

wobei:

$$(2.11)\quad \hat{\Xi} = (X_2, y^{\times})\,\Sigma^{-1}(I_n \beta)'\{(I_n \beta)\,\Sigma^{-1}(I_n \beta)'\}^{-1}.$$

Es gilt:

$$(2.12)\quad \left|(I_n \beta)\,\Sigma^{-1}(I_n \beta)'\right| = |\Sigma|^{-1}\,\begin{pmatrix}\beta\\ -1\end{pmatrix}'\Sigma\,\begin{pmatrix}\beta\\ -1\end{pmatrix}$$

(siehe Anhang 1.F2).

Weiterhin läßt sich zeigen, daß:

$$(2.13)\quad \text{sp}(X_2, y^{\times})\{\Sigma^{-1} - \Sigma^{-1}(I_n\beta)'\left[(I_n\beta)\Sigma^{-1}(I_n\beta)'\right]^{-1}(I_n\beta)\Sigma^{-1}\}(X_2, y^{\times})' = \frac{(y^{\times} - X_2\beta)'(y^{\times} - X_2\beta)}{\begin{pmatrix}\beta\\ -1\end{pmatrix}'\Sigma\,\begin{pmatrix}\beta\\ -1\end{pmatrix}}$$

(siehe Anhang 1.F4).

Es ergibt sich also aus den Eigenschaften der Normalverteilung:

$$(2.14)\quad p_1(\alpha,\beta,\Sigma \mid y, X_1, X_2) \propto \frac{p_o(\alpha,\beta,\Sigma)}{\left\{\begin{pmatrix}\beta\\ -1\end{pmatrix}'\Sigma\begin{pmatrix}\beta\\ -1\end{pmatrix}\right\}^{T/2}}\; e^{-\frac{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)}{2\begin{pmatrix}\beta\\ -1\end{pmatrix}'\Sigma\begin{pmatrix}\beta\\ -1\end{pmatrix}}}$$

Weitere Integration über ein Element von Σ stößt auf Schwierigkeiten. Eine andere Wahl der Parameter in der a priori Verteilung gestattet dagegen die Integration über $\sigma_{\omega\omega}$, die Varianz von ω, von 0 bis $+\infty$. Sei:

---

16) zur Begründung siehe z.B. A. Rényi (1962) S.158

(2.15) $\Sigma^{x} = \Sigma\sigma_{\omega\omega}^{-1}$

so läßt sich bei a priori Unkenntnis über $\sigma_{\omega\omega}$ formulieren:

(2.16) $p_o(\alpha,\beta,\Sigma^{x},\sigma_{\omega\omega}) \propto \sigma_{\omega\omega}^{-1}\, p_o(\alpha,\beta,\Sigma^{x})$.

Diese a priori Annahme erscheint auf den ersten Blick problematisch, wird doch damit Unabhängigkeit zwischen Σ^{x} und $\sigma_{\omega\omega}$ unterstellt. Im Falle der uninformativen a priori Dichtefunktion:

(2.17) $p_o(\alpha,\beta,\Sigma) \propto |\Sigma|^{-\frac{n+2}{2}}$ 17)

läßt sich aber $\sigma_{\omega\omega}^{-1}$ durch die bekannten Regeln der Transformation von Zufallsvariablen als Faktor von den anderen Parametern isolieren. Die Jakobische Determinante der Transformation von Σ in Σ^{x} und $\sigma_{\omega\omega}$ ist nämlich gleich

$\sigma_{\omega\omega}^{\frac{(n+2)(n+1)}{2}-1}$ da Σ $\binom{n+2}{2}$ verschiedene Elemente enthält, aber $\sigma_{\omega\omega}$ nicht transformiert wird. Damit ergibt sich aus (2.17):

$$p_o(\alpha,\beta,\Sigma^{x},\sigma_{\omega\omega}) \propto |\Sigma^{x}|^{-\frac{n+2}{2}}\, \sigma_{\omega\omega}^{-1}.$$

Mithin erscheint die Wahl der a priori Dichte in (2.16) für den Fall mangelnder a priori Information, der in diesem Abschnitt im Vordergrund der Untersuchung stehen soll, gerechtfertigt. Insbesondere ist diese Parametrisierung aber auch dann sinnvoll, wenn a priori Information über die Verhältnisse der Fehlervarianzen bzw. Kovarianzen zu der Varianz von ω vorliegt, wenn man also wenigstens ungefähre Vorstellungen von den Größenordnungen der Streuung der Beobachtungsfehler und des Fehlers in der Gleichung hat.

---

17) Für die Annahme einer solchen a priori Dichte hinsichtlich Σ sprechen dieselben Argumente, wie im Falle (2.8). Eine andere Begründung ist zu finden bei C.Villegas(1969)

Mit der a priori Annahme (2.16) und der Likelihoodfunktion erhalten wir nach der Integration über Ξ:

$$(2.18)\quad p_1(\alpha,\beta,\Sigma^{\times},\sigma_{\omega\omega}|y,X_1,X_2) \propto p_o(\alpha,\beta,\Sigma^{\times})\sigma_{\omega\omega}^{-\frac{T}{2}-1}\{\begin{pmatrix}\beta\\-1\end{pmatrix}'\Sigma^{\times}\begin{pmatrix}\beta\\-1\end{pmatrix}\}^{-\frac{T}{2}}$$

$$e^{-\frac{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)}{2\sigma_{\omega\omega}\begin{pmatrix}\beta\\-1\end{pmatrix}'\Sigma^{\times}\begin{pmatrix}\beta\\-1\end{pmatrix}}} .$$

Diese Funktion läßt sich weiter über $\sigma_{\omega\omega}$ integrieren (vgl. Anhang 2.2) und es ergibt sich:

$$(2.19)\quad p_1(\alpha,\beta,\Sigma^{\times}|y,X_1,X_2) \propto p_o(\alpha,\beta,\Sigma^{\times})\{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)\}^{-\frac{T}{2}} .$$

Damit erhalten wir als erstes Ergebnis, daß wir durch die Stichprobe hinsichtlich $\Sigma^{\times}$ keine Information erhalten, denn integrieren wir (2.19) über α und β, so bleibt als a posteriori Randdichte von $\Sigma^{\times}$ gerade die a priori Dichtefunktion der $\binom{n+2}{2}-1$ unbekannten Elemente von $\Sigma^{\times}$ übrig, sofern jedenfalls a priori α und β unabhängig von $\Sigma^{\times}$ verteilt sind, insbesondere auch, wenn $p_o(\alpha,\beta,\Sigma^{\times}) \propto p_o(\Sigma^{\times})$.

Fassen wir weiter zusammen:

$$(2.20)\qquad X = (X_1,X_2) \quad \text{und} \quad \gamma = \begin{pmatrix}\alpha\\\beta\end{pmatrix} ,$$

so läßt sich für den zweiten Faktor in der a posteriori Randdichte von α und β schreiben:

$$(y - X\gamma)'(y - X\gamma) = (\gamma-\hat{\gamma})'X'X(\gamma-\hat{\gamma}) + y'y - \hat{\gamma}'X'X\hat{\gamma}$$
$$- y'X\gamma-\gamma'X'y + \hat{\gamma}'X'X\gamma + \gamma'X'X\hat{\gamma}.$$

Wählen wir:

$$(2.21)\qquad \hat{\gamma} := (X'X)^{-1}X'y$$

sofern Rang(X) = m+n, so gilt weiter:

$$(y-X\gamma)'(y-X\gamma) = y'y - y'X(X'X)^{-1}X'y\{1 + \frac{(\gamma - \hat{\gamma})'X'X(\gamma - \hat{\gamma})}{y'y - y'X(X'X)^{-1}X'y}\}$$

und damit ist:

$$(2.22)\quad p_1(\gamma,\Sigma^{\times}|y,X) \propto p_0(\gamma,\Sigma^{\times})\ \{1+\frac{(\gamma-\hat{\gamma})'X'X(\gamma-\hat{\gamma})}{(T-m-n)\hat{\sigma}^2}\}^{-\frac{T}{2}}$$

$$(2.23)\quad \text{wobei}\quad \hat{\sigma}^2=\frac{1}{T-m-n}\ y'\{I-X(X'X)^{-1}X'\}y$$

die einfache Kleinstquadratschätzung der Varianz der Störvariablen der Regression von y auf X ist. Als weiteres Ergebnis kann also gefolgert werden, daß die a posteriori Verteilung der Regressionskoeffizienten ein Produkt aus der a priori Verteilung und der multivariaten Student-t Verteilung mit T-m-n Freiheitsgraden mit dem Mittelwert gleich der K-QSchätzung des Regressionskoeffizientenvektors und der Kovarianzmatrix gleich dem $\frac{T-m-n}{T-m-n-2}$ fachen der K-Q Schätzung der Kovarianzmatrix der Koeffizientenschätzungen ist, sofern die a priori Dichte über $\Sigma^{\times}$ integriert werden kann oder $\Sigma^{\times}$ gegeben ist und T-m-n >2 ist. Nimmt man weiter an, daß a priori die Koeffizienten voneinander unabhängig sind, so gelten die obigen Folgerungen den Eigenschaften der multivariaten t-Verteilung gemäß für jeden Teilvektor von γ, insbesondere also auch für β.

Erstaunlich ist natürlich, daß sich im Endeffekt der Fehler in den Beobachtungen von Ξ auf die Verteilung des Koeffizientenvektors nicht auswirkt insoweit als die Beobachtungswerte X_1 und X_2 in völlig gleichartiger Weise in die Formel (2.22) eingehen. Dies liegt allerdings mit an der a priori Dichte (2.16), während nach der zunächst getroffenen Wahl der Parameter in (2.8) in Gleichung (2.14) α nur im Zähler des Exponenten in gleicher Weise wie ß auftaucht.

2.2 Unabhängigkeit zwischen den Fehlern in der zu erklärenden Variablen einerseits und den Fehlern in den erklärenden Variablen andererseits

Da das in Abschnitt 2.1 erzielte Ergebnis in Widerspruch zu dem der "klassischen" Statistik zu stehen scheint sei nun untersucht, ob mit etwas mehr a priori Kenntnissen, nämlich mit der Annahme, daß die Fehler-bzw. Störvariable ω unabhängig von den Fehlern in den Messungen der erklärenden Variablen ist, die selbst untereinander abhängig

sein sollen, ein anderes Ergebnis erzielt werden kann. Es sei also vorausgestzt, daß der Kovarianzvektor zwischen v_t und ω_t: $\sigma_{v\omega} = 0$ ist.Diese Annahme wäre z.B. dann berechtigt, wenn die zu erklärende Variable exakt gemessen worden ist, denn der Fehler in der Gleichung kann sicherlich unabhängig von den Beobachtungsfehlern sein.

Die gemeinsame a posteriori Dichtefunktion unter Beibehaltung der übrigen Modellannahmen vereinfacht sich mit $\sigma_{v\omega} = 0$ zu:

$$(2.24)\quad p_1(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega},\Xi|y,X_1,X_2) \propto p_o(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega})\sigma_{\omega\omega}^{-\frac{T}{2}}|\Sigma_v|^{-\frac{T}{2}}$$
$$e^{-\frac{1}{2}\,sp\{(X_2-\Xi)\Sigma_v^{-1}(X_2-\Xi)'+\sigma_{\omega\omega}^{-1}(y-X_1\alpha-\Xi\beta)(y-X_1\alpha-\Xi\beta)'\}}$$

wobei Σ_v die Kovarianzmatrix der Fehler in den erklärenden Variablen ist.

Die Integration über Ξ ergibt (analog zu (2.14)):

$$(2.25)\quad p_1(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega}|y,X_1,X_2) \propto p_o(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega})(\sigma_{\omega\omega}+\beta'\Sigma_v\beta)^{-\frac{T}{2}}$$
$$e^{-\frac{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)}{2(\sigma_{\omega\omega}+\beta'\Sigma_v\beta)}}.$$

Auch (2.25) läßt sich mit einer a priori Dichte entsprechend (2.17) - also:

$$(2.26)\quad p_o(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega}) \propto |\Sigma_v|^{-\frac{n+1}{2}}\ \sigma_{\omega\omega}^{-1}\ -$$

nicht ohne Transformation der Parameter in Σ_v weiter über $\sigma_{\omega\omega}$ integrieren. Es läßt sich hier auch eine zu der im Abschnitt 2.1 völlig analoge Änderung der Parametrisierung durchführen und zwar sei:

$$(2.27)\quad \Sigma_v^{\times} := \Sigma_v\,\sigma_{\omega\omega}^{-1}\ .$$

Mit (2.27) ergibt sich aus (2.26) durch Berücksichtigung der Jakobischen Determinante (in Σ_v sind $\binom{n+1}{2}$ verschiedene Parameter enthalten):

$$(2.28)\quad p_o(\alpha,\beta,\Sigma_v^{\times},\sigma_{\omega\omega}) \propto |\Sigma_v^{\times}|^{-\frac{n+1}{2}} \sigma_{\omega\omega}^{-1} .$$

Mit dieser a priori Dichte läßt sich (2.25) wie folgt schreiben:

$$(2.29)\quad p_1(\alpha,\beta,\Sigma_v^{\times},\sigma_{\omega\omega}|y,X_1,X_2) \propto |\Sigma_v^{\times}|^{-\frac{n+1}{2}} \sigma_{\omega\omega}^{-\frac{T}{2}-1} (1+\beta'\Sigma_v^{\times}\beta)^{-\frac{T}{2}} e^{-\frac{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)}{2\,\sigma_{\omega\omega}(1+\beta'\Sigma_v^{\times}\beta)}} .$$

Die Integration über $\sigma_{\omega\omega}$ von 0 bis $+\infty$ (entsprechend Anhang 2.2) ergibt:

$$(2.30)\quad p_1(\gamma,\Sigma_v^{\times}|y,X) \propto |\Sigma_v^{\times}|^{-\frac{n+1}{2}} \{1+\frac{(\gamma-\hat{\gamma})'X'X(\gamma-\hat{\gamma})}{(T-m-n)\,\hat{\sigma}^2}\}^{-\frac{T}{2}}$$

wobei $\gamma, \hat{\gamma}, \hat{\sigma}$ und X wie in (2.20), (2.21) und (2.23) definiert sind.

Die a priori Kenntnis über $\sigma_{v\omega}$ ändert also letztlich unter Aufrechterhaltung der sonstigen Annahmen nichts an dem Beitrag der Stichprobe zur Information über die Regressionskoeffizienten. Dasselbe Ergebnis erhalten wir übrigens, wenn wir noch einen Schritt weitergehen und Unabhängigkeit zwischen allen Fehlern postulieren, also annehmen, daß Σ Diagonalgestalt hat.

Unbefriedigend an allen diesen Resultaten ist zweierlei: Einmal wissen wir, daß die einfache K-Q-Methode zu inkonsistenten Schätzungen für die Koeffizienten und die Varianz der Störvariablen u der Regression von y auf X_1 und X_2 führt, wenn

$$(2.31)\qquad \operatorname*{plim}_{T\to\infty} \frac{1}{T}\,\Xi'\Xi \;<\; \infty$$

und mit $T\to\infty$ konvergiert γ gegen den Mittelwert der a posteriori Dichte $\hat{\gamma}$.[18] Zum anderen ist die völlige Ignorierung des Fehlers in den Beobachtungen, wie sie in der gleichen Behandlung von α und β bzw. X_1 und X_2 in (2.22)

---

[18] In besimmten Fällen kann aber die K-Q Schätzung auch mit (2.31) konsistent sein. Vgl. D.J. Aigner (1973).

und (2.30) zum Ausdruck kommt, Anlaß zur Kritik. Beide Einwände hängen allerdings insoweit zusammen, als ohne Fehler in den Variablen natürlich α und β von vornherein zusammengefaßt werden könnten und die t-Verteilung entsprechend dem zweiten Faktor von (2.22) bzw. (2.30) bei fehlender a priori Information als a posteriori Randdichte für γ resultieren würde.[19] Als Ansatzpunkt einer Erklärung für das zunächst paradox scheinende Ergebnis muß die Spezifikation der a priori Dichte über Ξ angesehen werden, denn mit $p_o(\Xi) \propto 1$ ist gerade (2.31) nicht erfüllt, da im Falle dieser diffusen Verteilung keine Kovarianzmatrix existiert und damit auch die Stichprobenkovarianzmatrix nicht konvergiert. Mit der Annahme $\operatorname{plim}_{T\to\infty} \frac{1}{T} \Xi'\Xi = \infty$ kann aber die K-Q Schätzung für γ konsistent sein, sofern $\Sigma < \infty$, wenn wir also - drastisch ausgedrückt - annehmen, daß die ξ_i so stark streuen, daß die Beobachtungsfehler, die selbst eine endliche Varianz haben, keine Rolle mehr spielen.[20]

Aber noch in anderer Hinsicht sind Einschränkungen zu machen: Die Änderung der Parametrisierung zum Beispiel von Σ_v zu $\Sigma_v^{\times}$ im Abschnitt 2.2 führt nur im Falle der diffusen a priori Dichte über die Parameter zu den gewonnenen Ergebnissen, was letztlich daran liegt, daß nur im Grenzfall der diffusen a priori Verteilung die Unabhängigkeit von Σ_v und $\sigma_{\omega\omega}$ die Unabhängigkeit von $\Sigma_v^{\times}$ und $\sigma_{\omega\omega}$ impliziert. Wir haben es also nach einer solchen Änderung der Parameter und der Annahme der Unabhängigkeit der neuen Parameter voneinander eigentlich mit einem neuen Modell zu tun. Haben wir a priori eine echte Dichtefunktion über

---

19) vgl. A. Zellner und G.C.Tiao (1964)

20) Eine solche Annahme kann durchaus ökonomischen Gegebenheiten eher gerecht werden als (2.31) vgl. D.J. Aigner (1966) Der Nachweis der Konsistenz gelingt z.B. im Falle trendbehafteter Variablen, wenn der Trend nicht zu stark ist.

Σ_v und $\sigma_{\omega\omega}$ angenommen, so sind eben $\Sigma_v^{\times}$ und $\sigma_{\omega\omega}$ offensichtlich voneinander abhängig und es gelingt zwar noch durch diese Transformation der Parameter die Integration über $\sigma_{\omega\omega}$ aber die a posteriori Dichte von α, β und $\Sigma_v^{\times}$ ist nicht mehr in ein Produkt aus der Randdichte von α und β auf der einen und $\Sigma_v^{\times}$ auf der anderen Seite zu zerlegen, wie das in (2.30) der Fall ist. Würden wir dagegen von vornherein mit $\Sigma_v^{\times}$ arbeiten und eine echte a priori Dichte über $\Sigma_v^{\times}$ annehmen, so ließe sich das Ergebnis sofort aus (2.30) ablesen.

2.3 Einfache Regression

Da ich mich in den weiteren Untersuchungen hin und wieder auf den Fall einer einzigen erklärenden Variablen (also auf einfache Regression) beschränken möchte, seien zuvor noch zur leichteren Vergleichbarkeit das zugrunde liegende Modell und die im Falle multipler Regression gewonnenen Ergebnisse für den Sonderfall $m = n = 1$ hingeschrieben, wobei die Elemente von X_1 alle gleich 1 sein sollen.

(2.1) entspricht:

$$x = \xi + v \qquad (2.1^{\times})$$

und (2.5) geht über in:

$$y = \alpha + \xi\beta + \omega \qquad (2.6^{\times})$$

wobei ξ der T-Spaltenvektor der Realisationen,
x der T-Spaltenvektor der Beobachtungen
und v der T-Spaltenvektor der Fehler in den Beobachtungen der erklärenden Variablen ist. α und β sind in diesem einfachen Regressionsmodell Skalare, während y und ω genau wie im multiplen Modell zu verstehen sind. Die Annahmen (A1) und (A2) sollen mit $n = 1$ weiter gelten. Die Kovarianzmatrix der Fehler ist dann eine 2×2 Matrix:

$$\Sigma = \begin{pmatrix} \sigma_{vv} & \sigma_{v\omega} \\ \sigma_{v\omega} & \sigma_{\omega\omega} \end{pmatrix}$$

und die a priori Dichte, die den Mangel an a priori Information zum Ausdruck bringen soll - also (2.17) vereinfacht sich zu:

$$(2.17^{\times})\quad p_o(\alpha,\beta,\Sigma,\xi) \propto (\sigma_{vv}\sigma_{\omega\omega}-\sigma_{v\omega}^2)^{-\frac{3}{2}},$$

und zusammen mit der Likelihood ergibt sich als gemeinsame a posteriori Dichte für die unbekannten Parameter:

$$(2.9^{\times})\quad p_1(\alpha,\beta,\Sigma,\xi|y,x) \propto (\sigma_{vv}\sigma_{\omega\omega}-\sigma_{v\omega}^2)^{-\frac{T+3}{2}} e^{-\frac{a}{2(\sigma_{vv}\sigma_{\omega\omega}-\sigma_{v\omega}^2)}}$$

mit $a = \sigma_{\omega\omega}(x-\xi)'(x-\xi) - 2\sigma_{v\omega}(x-\xi)'(y-\iota\alpha-\xi\beta) + \sigma_{vv}(y-\iota\alpha-\xi\beta)'(y-\iota\alpha-\xi\beta)$

wobei $\iota' = (1,1,\ldots,1)$.

Die Integration über ξ liefert (gemäß Anhang 2.1):

$$(2.14^{\times})\quad p_1(\alpha,\beta,\Sigma|y,x) \propto (\sigma_{vv}\sigma_{\omega\omega}-\sigma_{v\omega}^2)^{-\frac{3}{2}} (\sigma_{\omega\omega}-2\sigma_{v\omega}\beta+\sigma_{vv}\beta^2)^{-\frac{T}{2}} e^{-\frac{(y-\iota\alpha-x\beta)'(y-\iota\alpha-x\beta)}{2(\sigma_{\omega\omega}-2\sigma_{v\omega}\beta+\sigma_{vv}\beta^2)}}.$$

Mit $\sigma_{vv}^{\times} = \frac{\sigma_{vv}}{\sigma_{\omega\omega}}$ und $\sigma_{v\omega}^{\times} = \frac{\sigma_{v\omega}}{\sigma_{\omega\omega}}$ läßt sich $(2.14^{\times})$ umformen zu:

$$(2.18^{\times})\quad p_1(\alpha,\beta,\sigma_{vv}^{\times},\sigma_{v\omega}^{\times},\sigma_{\omega\omega}|y,x) \propto \sigma_{\omega\omega}^{-\frac{T}{2}-1} (\sigma_{vv}^{\times}-\sigma_{v\omega}^{\times 2})^{-\frac{3}{2}} (1-2\sigma_{v\omega}^{\times}\beta+\sigma_{vv}^{\times}\beta^2)^{-\frac{T}{2}} e^{-\frac{(y-\iota\alpha-x\beta)'(y-\iota\alpha-x\beta)}{2\sigma_{\omega\omega}(1-2\sigma_{v\omega}^{\times}\beta+\sigma_{vv}^{\times}\beta^2)}}.$$

Integrieren wir weiter über $\sigma_{\omega\omega}$ (gemäß Anhang 2.2):

$$(2.22^{\times})\quad p_1(\alpha,\beta,\sigma_{vv}^{\times},\sigma_{v\omega}^{\times}|y,x) \propto (\sigma_{vv}^{\times}-\sigma_{v\omega}^{\times 2})^{-\frac{3}{2}} \{(y-\iota\alpha-x\beta)'(y-\iota\alpha-x\beta)\}^{-\frac{T}{2}}.$$

Der zweite Faktor von (2.22$^{\times}$) läßt sich noch in die Form einer bivariaten Student-t Dichtefunktion bringen:

$$(2.32)\quad p_1(\alpha,\beta,\sigma^{\times}_{vv},\sigma^{\times}_{v\omega}|y,x) \propto (\sigma^{\times}_{vv}-\sigma^{\times 2}_{v\omega})^{-\frac{3}{2}} \left[1+\frac{\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}' \begin{pmatrix} T & \Sigma x_t \\ \Sigma x_t & \Sigma x_t^2\end{pmatrix} \begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}}{(T-2)\,\hat{\sigma}^2}\right]^{-\frac{T}{2}}$$

wobei $\hat{\sigma}^2 = \frac{T}{T-2}\left(s_{yy} - \frac{s_{xy}^2}{s_{xx}}\right)$

mit $s_{xx} = \frac{1}{T}\{\sum_{t=1}^{T} x_t^2 - \frac{1}{T}(\sum_{t=1}^{T} x_t)^2\}$,

$s_{yy} = \frac{1}{T}\{\sum_{t=1}^{T} y_t^2 - \frac{1}{T}(\sum_{t=1}^{T} y_t)^2\}$ und

$s_{xy} = \frac{1}{T}\{\sum_{t=1}^{T} x_t y_t - \frac{1}{T}\sum_{t=1}^{T} x_t \sum_{t=1}^{T} y_t\}$

sowie $\hat{\alpha} = \dfrac{\sum_{t=1}^{T} x_t^2 \sum_{t=1}^{T} y_t - \sum_{t=1}^{T} x_t \sum_{t=1}^{T} x_t y_t}{T^2 s_{xx}}$

und $\hat{\beta} = \dfrac{s_{xy}}{s_{xx}}$.

Ohne Annahme einer echten a priori Dichte über $\sigma^{\times}_{vv}$ und $\sigma^{\times}_{v\omega}$ läßt sich also die Randverteilung von α und β nicht berechnen. Der zweite Faktor in (2.32) hat - wie erwähnt - die Form einer t-Dichtefunktion entsprechend (2.22). Hätten wir eine echte a priori Verteilung über $\sigma^{\times}_{vv}$ und $\sigma^{\times}_{v\omega}$ angenommen, so erhielten wir nach den Eigenschaften der Student-Verteilung als Randdichte von α:

$$(2.33)\quad p_1(\alpha|y,x) \propto \left[1+\frac{(\alpha-\hat{\alpha})^2}{(T-2)\,\hat{\sigma}^2\,\dfrac{\sum_{t=1}^{T} x_t^2}{T^2 s_{xx}}}\right]^{-\frac{T-1}{2}}$$

und als Randdichte von ß:

$$(2.34)\quad p_1(\beta|y,x) \propto \left[1 + \frac{(\beta-\hat{\beta})^2}{(T-2)\hat{\sigma}^2 \frac{1}{T\, s_{xx}}}\right]^{-\frac{T-1}{2}} .$$

Schauen wir uns noch rasch die Ergebnisse bei Unabhängigkeit von v und ω an. Das ist der von Zellner [21] untersuchte Fall und insbesondere ein Spezialfall des in Abschnitt 2.2 betrachteten Modells. Die gemeinsame a posteriori Dichte vereinfacht sich mit

$$p_o(\alpha,\beta,\sigma^{\times}_{vv},\sigma_{\omega\omega},\xi) \propto \frac{p_o(\sigma_{vv})}{\sigma_{\omega\omega}} \quad \text{wobei} \quad \sigma^{\times}_{vv} = \frac{\sigma_{vv}}{\sigma_{\omega\omega}}$$

zu:

$$(2.24^{\times})\quad p_1(\alpha,\beta,\sigma^{\times}_{vv},\sigma_{\omega\omega},\Xi|x,y) \propto p_o(\sigma^{\times}_{vv})\sigma^{\times-\frac{T}{2}}_{vv}\,\sigma_{\omega\omega}^{-\frac{T}{2}-1} e^{-\frac{(x-\xi)'(x-\xi)+\sigma^{\times}_{vv}(y-\iota\alpha-\xi\beta)'(y-\iota\alpha-\xi\beta)}{2\sigma_{\omega\omega}\sigma^{\times}_{vv}}}$$

Die Integration über die Elemente von ξ (gemäß Anhang 2.1) und über $\sigma_{\omega\omega}$ (gemäß Anhang 2.2) ergibt:

$$(2.30^{\times})\quad p_1(\alpha,\beta,\sigma^{\times}_{vv}|y,x) \propto p_o(\sigma^{\times}_{vv})\left[1 + \frac{\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}'\begin{pmatrix} T & \Sigma x_t\\ \Sigma x_t & \Sigma x_t^2\end{pmatrix}\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}}{(T-2)\,\hat{\sigma}^2}\right]^{-\frac{T}{2}}$$

wobei $\hat{\alpha}$, $\hat{\beta}$ und $\hat{\sigma}$ genau wie im Anschluß an (2.32) definiert sind.

Natürlich bestätigt sich auch hier das Ergebnis der Betrachtung des entsprechenden multiplen Modells. Wiederum sieht man, daß in bezug auf das Verhältnis der Varianzen aus der Stichprobe keine Information gewonnen wird, sondern nach Integration von (2.30×) über α und ß gerade die a priori Verteilung für $\sigma^{\times}_{vv}$ übrigbleibt.

---

21) A. Zellner (1971) S. 132 f.

3 Echtes a priori Wissen

Wie in den einleitenden Erörterungen betont, liegt der eigentliche Grund für die Anwendung des Bayesschen Theorems in der methodischen Einbeziehung von zwar unvollständiger aber eben doch echter a priori Information.

3.1 A priori Information über die Regressionskoeffizienten

Verhältnismäßig leicht läßt sich a priori Information über den hier hauptsächlich interessierenden Parametervektor γ selbst in Form einer echten a priori Dichtefunktion - echt hinsichtlich γ - in das Modell einfügen. Es erscheint sinnvoll, dafür den Verteilungstyp zu wählen, der sich als a posteriori Verteilung für den Koeffizientenvektor γ bei a priori Unkenntnis ergibt, also eine multivariate Student-t-Verteilung:

$$(3.1)\quad p_o(\gamma|T_o,\hat{\gamma}_o,\hat{\Sigma}_o) \propto \{1 + \frac{(\gamma-\hat{\gamma}_o)'\,\hat{\Sigma}_o^{-1}(\gamma-\hat{\gamma}_o)}{(T_o-m-n)}\}^{-\frac{T_o}{2}}$$

wobei T_o ein vorgegebener Skalar, $\hat{\gamma}_o$ ein gegebener $m+n$ Spaltenvektor und $\hat{\Sigma}_o$ eine gegebene symmetrische und positiv definite $(m+n)\times(m+n)$ Matrix sein soll. Es läßt sich leicht eine auch von subjektiver Introspektion skeptisch gegenüberstehenden Anhängern der "klassischen" Statistik annehmbare Begründung für diese a priori Wahrscheinlichkeitsverteilung geben, kann man sie doch als Information aus früher durchgeführten Untersuchungen des Zusammenhangs zwischen den betrachteten Variablen ansehen. (3.1) ist so formuliert, daß man einerseits den Zusammenhang zu (2.22) bzw. (2.30) sieht und andererseits sofort erkennt, daß man für die Konstanten, die die a priori Dichtefunktion bestimmen, die in regressionsanalytischen Untersuchungen üblicherweise angegebenen Schätzungen der Koeffizienten und der Kovarianzmatrix dieser Koeffizientenschätzungen sowie den Stichprobenumfang einsetzen kann. Es muß dabei allerdings gewährleistet sein, daß die in der zu untersuchenden Stichprobe vorliegenden Werte nicht schon zu diesen Schätzungen verwendet wurden. Dies wäre, wenn wir an makroökonomische Untersuchungen denken, z.B.

sichergestellt, wenn die in die a priori Dichtefunktion eingehenden Schätzungen aus einer Untersuchung derselben Regression in einer anderen - möglichst vergleichbaren - Volkswirtschaft stammen. Natürlich wäre kaum anzunehmen, daß beispielsweise die marginale Konsumquote in den USA exakt dieselbe wäre wie in der BRD und daß die Fehler in den Beobachtungen hier wie dort dieselbe Streuung haben, dennoch werden die Ergebnisse von Untersuchungen über die Grenzneigugg zum Konsum in den USA unseren Informationsstand über die gleiche Größe in der BRD verbessern, denn so sehr groß werden wir die Abweichung nicht erwarten. Natürlich kann man auch (3.1) direkt als Ausdruck subjektiver Glaubwürdigkeitsvorstellungen ansehen, in denen frühere empirische Arbeiten aber auch theoretische Erwägungen - etwa derart, daß eine Konsumquote größer als eins für wenig wahrscheinlich gehalten wird - ihren Niederschlag gefungen haben.

Gehen wir davon aus, daß $\sigma_{v_\omega} = 0$ und

$$(3.2)\quad p_o(\gamma, \Sigma_v^{\times}, \sigma_{\omega\omega}, \Xi \mid T_o, \hat{\gamma}_o, \hat{\Sigma}_o) \propto \frac{p_o(\gamma \mid T_o, \hat{\gamma}_o, \hat{\Sigma}_o)\, p_o(\Sigma_v^{\times})}{\sigma_{\omega\omega}}$$

so ergibt sich a posteriori nach Integration über Ξ und $\sigma_{\omega\omega}$ und $\Sigma_v^{\times}$

$$(3.3)\quad p_o(\gamma \mid T_o, \hat{\gamma}_o, \hat{\Sigma}_o, y, X) \propto \left(1 + \frac{(\gamma - \hat{\gamma}_o)' \hat{\Sigma}_o^{-1} (\gamma - \hat{\gamma}_o)}{(T_o - m - n)}\right)^{-\frac{T}{2}_o} \left(1 + \frac{(\gamma - \hat{\gamma})' X' X (\gamma - \hat{\gamma})}{(T - m - n)\, \hat{\sigma}^2}\right)^{-\frac{T}{2}}$$

wobei $\hat{\gamma}, \hat{\sigma}^2$ und X in (2.20), (2.21) und (2.23) definiert sind.

Wir erhalten also als Randdichte für γ a posteriori eine multivariate "Doppel-t-Verteilung", von A. Zellner und G.C. Tiao so genannt, weil diese Dichtefunktion aus zwei Faktoren besteht, von denen jeder für sich - abgesehen von der Normierungskonstanten - die Form einer multivariaten t-Dichte hat.[22)]

Allerdings ist nur der zweite Faktor in (3.3), in dem die Kleinstquadratschätzungen für γ und die Varianz der γ-Schätzung die entscheidenden Rollen spielen, vom Stichprobenumfang abhängig. Letztlich bedeutet dies, daß mit zunehmendem Stichprobenumfang der Modalwert der a posteriori Verteilung gegen die K-Q-Schätzung von γ konvergiert. Dies gilt übrigens unabhängig von dem a priori Verteilungstyp für γ, sofern wir nicht gerade eine Verteilung wählen, die nur auf einem endlichen Intervall der reellen Zahlen positive Dichte hat und solange wir an den a priori Annahmen hinsichtlich Ξ und Σ nichts ändern. Sobald wir nämlich annehmen, daß a priori:

$$p_o(\gamma, \Sigma_v^x, \sigma_{\omega\omega}, \Xi) \propto p_o(\gamma)\; p_o(\Sigma_v^x)\; \sigma_{\omega\omega}^{-1}$$

ist, erhalten wir als a posteriori Verteilung ein Produkt aus a priori Verteilungen einerseits und der multivariaten t-Verteilung andererseits, wobei nur dieser letzte Faktor von der Stichprobe abhängt, d.h. an der Schlüsselstellung der K-Q-Schätzungen wird sich nichts ändern. Ein anderes Ergebnis ist nur zu erwarten, wenn wir a priori Information über Σ oder Ξ unterstellen.

3.2 A priori Information über die Kovarianzmatrix der Fehler

Nehmen wir hinsichtlich Ξ weiter eine Quasi-Gleichverteilung an und ist Σ exakt bekannt, so erhalten wir als a posteriori Randdichtefunktion für die Regressionskoeffizienten (entsprechend (2.14)):

$$(3.4)\quad p_1(\alpha, \beta | \Sigma, y, X_1, X_2) \propto \frac{p_o(\alpha, \beta | \Sigma)}{\{(\beta_{-1})' \Sigma (\beta_{-1})\}^{\frac{T}{2}}}\; e^{-\frac{(y - X_1\alpha - X_2\beta)'(y - X_1\alpha - X_2\beta)}{2(\beta_{-1})' \Sigma (\beta_{-1})}}.$$

---

22) Diese Verteilung wurde in anderem Zusammenhang von G.C. Tiao und A. Zellner (1964) formuliert und näher untersucht, siehe auch: A. Zellner (1971), S. 101 und 110f.

Nehmen wir weiter an, daß

(3.5) $p_o(\alpha, \beta|\Sigma) \propto 1$

so ergibt die Integration über α (nach Anhang 2.1[x]):

$$p_1(\beta|\Sigma, y, X_1, X_2) \propto \{ \begin{pmatrix} \beta \\ -1 \end{pmatrix}' \Sigma \begin{pmatrix} \beta \\ -1 \end{pmatrix} \}^{-\frac{T-m}{2}} e^{-\frac{(y-X_2\beta)'P_1(y-X_2\beta)}{2\begin{pmatrix} \beta \\ -1 \end{pmatrix}' \Sigma \begin{pmatrix} \beta \\ -1 \end{pmatrix}}}$$

wobei:

(3.6) $P_1 = I - X_1(X_1'X_1)^{-1}X_1'$.

Der Exponent läßt sich umformen, so daß:

$$(3.7)\quad p_1(\beta|\Sigma, y, X_1, X_2) \propto \{ \begin{pmatrix} \beta \\ -1 \end{pmatrix}' \Sigma \begin{pmatrix} \beta \\ -1 \end{pmatrix} \}^{-\frac{T-m}{2}} e^{-\frac{(\beta-\hat{\beta})'X_2'P_1X_2(\beta-\hat{\beta}) + y'P_1P_{21}^{-1}P_1y}{2\begin{pmatrix} \beta \\ -1 \end{pmatrix}' \Sigma \begin{pmatrix} \beta \\ -1 \end{pmatrix}}}$$

wobei

(3.8) $P_{21} = I - X_2(X_2'P_1X_2)^{-1}X_2'$

und

(3.9) $\hat{\beta} = (X_2'P_1X_2)^{-1}X_2'P_1y$.

$\hat{\beta}$ ist also nichts anderes als die K-Q-Schätzung der Koeffizienten β der Regression von y auf X_1 und X_2.

Die resultierende a posteriori Verteilung von β ist leider nicht von bekanntem Typ. Die einfache K-Q-Schätzung spielt zwar weiterhin eine Rolle, aber die Information über β führt doch zu einschneidenden Veränderungen, z.B. auch des Modalwertes der a posteriori Verteilung. In dem Falle der einfachen Regression ($m = n = 1$ und X_1 Scheinvariable) erhalten wir als Bestimmungsgleichung für den Modalwert:

(3.10) $c_1\beta^3 + c_2\beta^2 + c_3\beta + c_4 = 0$

wobei

$c_1 = (T-1)\sigma_{vv}^2$

$c_2 = -3(T-1)\sigma_{vv}\sigma_{v\omega} + Ts_{xx}(\sigma_{vv}\hat{\beta} - \sigma_{v\omega})$

$$c_3 = (T-1)(\sigma_{vv}\sigma_{\omega\omega} + \sigma_{v\omega}^2) + Ts_{xx}\{\sigma_{\omega\omega} - \sigma_{vv}(\hat{\beta}^2 + (T-2)\hat{\sigma}_{\hat{\beta}}^2)\}$$

und

$$c_4 = -(T-1)\sigma_{v\omega}\sigma_{\omega\omega} + Ts_{xx}\{\sigma_{v\omega}(\hat{\beta}^2 + (T-2)\hat{\sigma}_{\hat{\beta}}^2) - \sigma_{\omega\omega}\hat{\beta}\}$$

mit $\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{s_{xx}}$.

Exakte Information über Σ wird nun allerdings in praktischen Fällen selten zur Verfügung stehen. Gewöhnlich steht zumindest die Varianz des Fehlers in der Gleichung nach den Regressionskoeffizienten im Mittelpunkt der statistischen Auswertung der Daten. Nun liegt -wie erwähnt - der Vorzug der Bayesschen Analyse in der systematischen Einbeziehung unvollständiger a priori Information, die allerdings in die Form einer a priori Wahrscheinlichkeitsverteilung gegossen werden muß.

Es verlangen in diesem Zusammenhang zwei Fälle unterschiedliche Behandlung:

a) Es liegen Informationen nur über einzelne Bestandteile der Kovarianzmatrix vor.
b) Es sind Informationen über die gesamte Kovarianzmatrix in einer a priori Dichtefunktion enthalten.

Der erste Fall betrifft insbesondere Modelle, in denen zwischen den Beobachtungsfehlern der erklärenden Variablen und ω Unabhängigkeit besteht, also $\sigma_{v\omega} = 0$ ist, denn damit ist bereits ein Teil von Σ bekannt. Wollen wir nun a priori Information über $\sigma_{\omega\omega}$ einführen, so erhebt sich die Frage, welche Verteilung zur Aufnahme dieses a priori Wissens geeignet erscheint. Ein sich als brauchbar erweisender Dichtefunktionstyp zur Einbeziehung von a priori Information über eine Varianz im Zusammenhang mit Normalverteilungsvoraussetzungen ist die Klasse der inversen χ^2 Verteilungen[23], denn genau eine solche Verteilung ergibt sich als bedingte a posteriori Verteilung für $\sigma_{\omega\omega}$ bei gegebenem α, β und Σ_v^x, wenn man a priori davon ausgeht, daß $\log\sigma_{\omega\omega}$ und Ξ diffus gleichverteilt sind (vgl. (2.29)). Sei also angenommen, daß:

$$(3.13)\quad p_o(\alpha,\beta,\Sigma_v^{\times},\sigma_{\omega\omega},\Xi \mid s,k) \propto p_o(\Sigma_v^{\times})\,\sigma_{\omega\omega}^{-\frac{k}{2}-1}\, e^{-\frac{s}{2\sigma_{\omega\omega}}}$$

wobei s und k positive Konstanten sind. Mit dieser a priori Dichte erhalten wir zusammen mit der Likelihood-funktion nach Integration über Ξ:

$$(3.14)\quad p_1(\alpha,\beta,\Sigma_v^{\times},\sigma_{\omega\omega} \mid y,X_1,X_2,s,k) \propto p_o(\Sigma_v^{\times})\,(1+\beta'\Sigma_v^{\times}\beta)^{-\frac{T}{2}}$$

$$\sigma_{\omega\omega}^{-\frac{T+k}{2}-1}\, e^{-\frac{(1+\beta'\Sigma_v^{\times}\beta)s+u'u}{2(1+\beta'\Sigma_v^{\times}\beta)\sigma_{\omega\omega}}}$$

wobei

$$(3.15)\quad u := y - X_1\alpha - X_2\beta.$$

Die Integration über $\sigma_{\omega\omega}$ ergibt (entsprechend Anhang 2.2):

$$p_1(\alpha,\beta,\Sigma_v^{\times} \mid y,X_1,X_2,s,k) \propto p_o(\Sigma_v^{\times})\,(1+\beta'\Sigma_v^{\times}\beta)^{\frac{k}{2}}$$

$$\{(1+\beta'\Sigma_v^{\times}\beta)s+u'u\}^{-\frac{T+k}{2}}.$$

Dies läßt sich umformen zu:

$$(3.16)\quad p_1(\alpha,\beta,\Sigma_v^{\times} \mid y,X_1,X_2,s,k) \propto p_o(\Sigma_v^{\times})\,(1+\beta'\Sigma_v^{\times}\beta)^{\frac{k}{2}}$$

$$\left[(T-m-n)\hat{\sigma}^2+(1+\beta'\Sigma_v^{\times}\beta)s+\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}'\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}\right]^{-\frac{T+k}{2}}$$

wobei $\hat{\beta}$ in (3.9) und $\hat{\sigma}^2$ in (2.23) definiert sind und

---

23) Siehe auch A.Zellner(1971)S.136 und Florens, Mouchart und Richard (1972) S.17. Zur Erläuterung der Bezeichnung: Ist Z χ^2 verteilt mit k Freiheitsgraden, also $f_Z(z) \propto z^{k/2-1}e^{-sz/2}$ für $z \in R_+$ und $s>0$, so heißt $\Sigma = Z^{-1}$ invers χ^2 verteilt mit k Freiheitsgraden. Die Dichte der inversen χ^2_k Verteilung lautet also:
$f_\Sigma(\sigma) \propto \sigma^{-\frac{k}{2}-1}\, e^{-\frac{s}{2\sigma}}$ für $\sigma \in R_+$.

(3.17) $\hat{\alpha} := (X_1' P_2 X_1)^{-1} X_1' P_2 y$

mit

(3.18) $P_2 := I - X_2 (X_2' X_2)^{-1} X_2'$

ist. Integrieren wir noch die zu exakt gemessenen Variablen gehörenden Koeffizienten heraus (im Falle n = 1 also das konstante Glied), so ergibt sich entsprechend Anhang 2.3$^{\times}$:

(3.19) $$p_1(\beta, \Sigma_v^{\times} | y, X_1, X_2, s, k) \propto p_0(\Sigma_v^{\times}) (1 + \beta' \Sigma_v^{\times} \beta)^{\frac{k}{2}}$$
$$\{(T-m-n)\hat{\sigma}^2 + (1+\beta' \Sigma_v^{\times} \beta) s + (\beta - \hat{\beta})' X_2' P_1 X_2 (\beta - \hat{\beta})\}^{-\frac{T+k-m}{2}}$$

wobei P_1 in (3.6) defininiert ist.

Bei gegebenem $\Sigma_v^{\times}$ erhalten wir also als bedingte a posteriori Verteilung für β keine t-Verteilung mehr. Immerhin läßt sich die Dichte dieser bedingten Verteilung aber wie folgt schreiben:

(3.20) $$p_1(\beta | \Sigma_v^{\times}, s, k, y, X_1 X_2) \propto (1 + \beta' \Sigma_v^{\times} \beta)^{\frac{k}{2}}$$
$$\left(1 + \frac{(\beta - \hat{\beta}_s)'(s_v + X_2' P_1 X_2)(\beta - \hat{\beta}_s)}{s + (T-m-n)\hat{\sigma}^2 + \hat{\beta}' X_2' P_1 [I - X_2 (s\Sigma_v^{\times} + X_2' P_1 X_2)^{-1} X_2' P_1 X_2 \hat{\beta}}\right)^{-\frac{T+k-m}{2}}$$

wobei

(3.21) $\hat{\beta}_s := (s\Sigma_v^{\times} + X_2' P_1 X_2)^{-1} X_2' P_1 X_2 \hat{\beta}$.

Der zweite Faktor in (3.20) hat die Form einer multivariaten t-Dichte mit dem Mittelwert $\hat{\beta}_s$. Im Falle der einfachen Regression können wir folgern, daß $\hat{\beta}_s$ dem Betrage nach kleiner als $\hat{\beta}$ ist, das bekanntlich unter der Annahme plim $\frac{1}{T}\xi'\xi < \infty$ den Koeffizienten β bereits dem Betrage nach unterschätzt. Auch im multiplen Fall wäre zur Erreichung einer konsistenten Schätzung für β im ersten Faktor von (3.21) $s = -T\sigma_{\omega\omega}$ zu wählen, was hier

leider ausgeschlossen ist, da die a priori Verteilung (3.13) nur für $s > 0$ definiert ist. Mit zunehmendem Stichprobenumfang wird aber $X_2' P_1 X_2$ größer und es gilt:

$$\operatorname*{plim}_{T\to\infty} \hat{\beta}_s = \operatorname*{plim}_{T\to\infty} \left(\frac{s\Sigma_v^{\times}}{T} + \frac{X_2' P_1 X_2}{T} \right)^{-1} \frac{X_2' P_1 X_2}{T} \hat{\beta} \quad \Rightarrow$$

$$\text{(3.22)} \quad \operatorname*{plim}_{T\to\infty} \hat{\beta}_s = \operatorname*{plim}_{T\to\infty} \hat{\beta}$$

und ebenso gilt für die Varianzschätzung von $\hat{\beta}_s$:

$$\text{(3.23)} \quad \operatorname*{plim}_{T\to\infty} \frac{s\Sigma_v^{\times} + X_2' P_1 X_2}{s+(T-m-n)\hat{\sigma}^2+\hat{\beta}' X_2' P_1 \{I - X_2 (s\Sigma_v^{\times}+X_2' P_1 X_2)^{-1} X_2'\} P_1 X_2 \hat{\beta}}$$

$$= \operatorname*{plim}_{T\to\infty} \frac{X_2' P_1 X_2}{(T-m-n)\hat{\sigma}^2} \quad .$$

Da der erste Faktor in (3.20) nicht vom Stichprobenumfang abhängt, bleibt auch bei dieser a priori Information die Modifikation der K-Q-Schätzungen auf kleine Stichproben beschränkt.

Die gewonnenen Ergebnisse enthalten auch den Grenzfall fehlender a priori Kenntnis über $\sigma_{\omega\omega}$. (3.13) geht nämlich in die diffuse Verteilung

$$p_o(\alpha, \beta, \Sigma_v^{\times}, \sigma_{\omega\omega}, \Xi) \propto p_o(\Sigma_v^{\times})\ \sigma_{\omega\omega}^{-1}$$

über, wenn $k = 0$ und $s = 0$. In diesem Fall entspricht (3.16) genau (2.30), und (3.19) mit $k = 0$ und $s = 0$ können wir auch erhalten, indem wir (2.30) über α integrieren.

Bleibt der Fall zu betrachten, in dem unvollständige Informationen über alle Komponenten der Kovarianzmatrix der Fehler gemeinsam vorliegen. Geeignet zur Einbeziehung derartigen a priori Wissens in das betrachtete Modell erscheint die Klasse der inversen Wishart-Verteilungen. Eine Verteilung dieses Typs erhalten wir z.B. indem wir aus der a posteriori Dichte (2.9) die bedingte

Dichte für Σ bei gegebenem α, β und Ξ und diffuser a priori Verteilung über Σ - genauer $p_o(\Sigma) \propto |\Sigma|^{-\frac{n+2}{2}}$ - berechnen.[24] Nehmen wir also als a priori Wahrscheinlichkeitsdichtefunktion:

$$(3.24)\quad p_o(\alpha,\beta,\Xi,\Sigma|S,k) \propto p_o(\alpha,\beta)\,\frac{|S|^{\frac{k}{2}}}{|\Sigma|^{\frac{k+n+2}{2}}}\,e^{-\frac{1}{2}\,sp(\Sigma^{-1}S)}$$

wobei S eine vorgegebene symmetrische und positiv definite $(n+1)\times(n+1)$ Matrix und k ein positiver Skalar sein soll $(k \geq n+1)$.

Multipliziert mit der Likelihoodfunktion (2.7) erhalten wir a posteriori:

$$(3.25)\quad p_1(\alpha,\beta,\Xi,\Sigma|S,k,y,X_1,X_2) \propto \frac{p_o(\alpha,\beta)}{|\Sigma|^{\frac{T+k+n+2}{2}}}\,e^{-\frac{1}{2}\,sp(\Sigma^{-1}B)}$$

wobei $B := S + (X_2 - \Xi, y - X_1\alpha - \Xi\beta)'(X_2 - \Xi, y - X_1\alpha - \Xi\beta)$.

In diesem Fall erscheint es zweckmäßiger, zuerst über die Kovarianzmatrix Σ anstatt über Ξ zu integrieren, um die Randdichtefunktionen der Koeffizienten α und β zu errechnen, denn nach (3.24) ist die bedingte a

---

[24] Diesen Verteilungstyp schlagen auch Florens, Mouchart und Richard (1972) S.20f. vor. Zur Bezeichnung: Ist die symmetrische und positiv definite $r\times r$ Matrix Z Wishart verteilt mit den Parametern C und k, also

$$f_Z(Z|C,k) \propto |Z|^{\frac{k+r-1}{2}}\,|C|^{-\frac{k}{2}}\,e^{-\frac{1}{2}\,sp(C^{-1}Z)}$$

wobei $r \leq k$ und C eine symmetrische und positiv definite $r\times r$ Matrix ist, so heißt $\Sigma := Z^{-1}$ invers Wishart verteilt mit den Parametern C^{-1} und k. Die Dichtefunktion der inversen Wishart-Verteilung lautet:

$$f_\Sigma(\Sigma|C^{-1},k) \propto |\Sigma|^{-\frac{k+r+1}{2}}\,|C|^{-\frac{k}{2}}\,e^{-\frac{1}{2}\,sp\,(C^{-1}\,\Sigma^{-1})}$$

posteriori Verteilung von Σ bei gegebenem α, β und Ξ auch aus der Klasse der inversen Wishart-Verteilungen:

$$(3.26)\quad p_1(\Sigma|\alpha,\beta,\Xi,D) \propto \frac{|B|^{\frac{T+k}{2}}}{|\Sigma|^{\frac{T+k+n+2}{2}}} e^{-\frac{1}{2}\,sp(\Sigma^{-1}B)}$$

wobei D die Parameter der a priori Verteilung und die Daten zusammenfaßt. Damit läßt sich das Ergebnis der Integration von (3.25) über Σ leicht angeben:

$$p_1(\alpha,\beta,\Xi|D) \propto p_o(\alpha,\beta)\,|S + (X_2-\Xi,y-X_1\alpha-\Xi\beta)'(X_2-\Xi,y-X_1\alpha-\Xi\beta)|^{-\frac{T+k}{2}}.$$

Das läßt sich umformen zu (siehe Anhang 1, F1):

$$(3.27)\quad p_1(\alpha,\beta,\Xi|D) \propto p_o(\alpha,\beta)\,|I+(X_2-\Xi,y-X_1\alpha-\Xi\beta)S^{-1}(X_2-\Xi,y-X_1\alpha-\Xi\beta)'|^{-\frac{T+k}{2}}.$$

Über Ξ integriert ergibt sich (vgl. Anhang 2.3[xx]):

$$(3.28)\quad p_1(\alpha,\beta|D) \propto p_o(\alpha,\beta)\{\binom{\beta}{-1}'S\binom{\beta}{-1}\}^{\frac{k-n}{2}}\{\binom{\beta}{-1}'S\binom{\beta}{-1} + (T-m-n)\hat{\sigma}^2 + \begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}'\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}\}^{-\frac{T+k-n}{2}}.$$

Mit $p_o(\alpha,\beta) \propto 1$ ist aus (3.28) zu folgern, daß die bedingte a posteriori Verteilung von α bei gegebenem β aus der Klasse der multivariaten t-Verteilungen stammt. (3.28) läßt sich demnach noch über α integrieren. Es ergibt sich (gemäß Anhang 2.3[x]):

$$(3.29)\quad p_1(\beta|D) \propto \{\binom{\beta}{-1}'S\binom{\beta}{-1}\}^{\frac{k-n}{2}}\{(T-m-n)\hat{\sigma}^2 + \binom{\beta}{-1}'S\binom{\beta}{-1} + (\beta-\hat{\beta})'X_2'P_1X_2(\beta-\hat{\beta})\}^{-\frac{T+k-n-m}{2}}$$

was sich bei Unterteilung von S in: $S = \begin{pmatrix}S_v & s_{v\omega}\\ s_{v\omega}' & s_{\omega\omega}\end{pmatrix}$

wie folgt schreiben läßt:

$$(3.30)\quad p_1(\beta|D) \propto \{1 + \frac{(\beta - S_v^{-1}s_{v\omega})' S_v (\beta - S_v^{-1}s_{v\omega})}{s_{\omega\omega} - s_{v\omega}' S_v^{-1} s_{v\omega}}\}^{\frac{k-n}{2}}$$

$$\{1 + \frac{(\beta - \hat{\beta}_S)'(S_v + X_2' P_1 X_2)(\beta - \hat{\beta}_S)}{(T-m-n)\hat{\sigma}_S^2}\}^{-\frac{T+k-n-m}{2}}$$

wobei

$$(3.31)\quad \hat{\beta}_S := (S_v + X_2' P_1 X_2)^{-1}(s_{v\omega} + X_2' P_1 X_2 \hat{\beta})$$

und

$$(3.32)\quad (T-m-n)\hat{\sigma}_S^2 := (T-m-n)\hat{\sigma}^2 + s_{\omega\omega} + \hat{\beta}' X_2' P_1 X_2 \hat{\beta}$$
$$- (s_{v\omega} + X_2' P_1 X_2 \hat{\beta})'(S_v + X_2' P_1 X_2)^{-1}(s_{v\omega} + X_2' P_1 X_2 \hat{\beta}).$$

Mit der a priori Dichtefunktion (3.24) erhalten wir also im Falle vagen a priori Wissens über α, β und Ξ als a posteriori Randdichte von β eine multivariate Verteilung, die man als verallgemeinerte Doppel-t-Verteilung bezeichnen könnte; verallgemeinert, weil der erste Faktor einen positiven Exponenten hat, wenn (3.24) eine nicht entartete inverse Wishart-Verteilung sein soll.

Das erscheint insbesondere für den Spezialfall der einfachen Regression interessant. Aus (3.30) ergibt sich mit $n = 1$, $X_1 = \iota$ und $X_2 = x$:

$$(3.30^{\times})\quad p_1(\beta|D) \propto \left(1 + \frac{(\beta - \frac{s_{v\omega}}{s_{vv}})^2}{\frac{s_{\omega\omega}s_{vv} - s_{v\omega}^2}{s_{vv}}}\right)^{\frac{k-1}{2}} \left(1 + \frac{(\beta - \hat{\beta}_S)^2}{\frac{(T-2)\hat{\sigma}_S^2}{s_{vv} + Ts_{xx}}}\right)^{-\frac{T+k-2}{2}}$$

wobei

$$(3.31^{\times})\quad \hat{\beta}_S := \frac{s_{v\omega} + Ts_{xy}}{s_{vv} + Ts_{xx}}$$

und

$$(3.32^{\times})\quad (T-2)\hat{\sigma}_S^2 := s_{\omega\omega} + Ts_{yy} - \frac{(s_{v\omega} + Ts_{xy})^2}{s_{vv} + Ts_{xx}}.$$

s_{xx}, s_{xy} und s_{yy} sind im Zusammenhang mit (2.32) definiert.

Wie man leicht sieht, ist:

$$\text{(3.33}^{\times}\text{)} \quad \operatorname*{plim}_{T\to\infty} \hat{\beta}_S = \operatorname*{plim}_{T\to\infty} \hat{\beta}$$

und

$$\text{(3.34}^{\times}\text{)} \quad \operatorname*{plim}_{T\to\infty} \hat{\sigma}_S^2 = \operatorname*{plim}_{T\to\infty} \hat{\sigma}^2$$

wobei $\hat{\beta}$ und $\hat{\sigma}$ die K-Q-Schätzungen des Koeffizienten β und der Varianz von $u := y - \iota\alpha - x\beta$ sind.

Die gleichen Ergebnisse erhalten wir im multiplen Fall:

$$\operatorname*{plim}_{T\to\infty} \hat{\beta}_S = \operatorname*{plim}_{T\to\infty} \left(\frac{1}{T} S_v + \frac{1}{T} X_2' P_1 X_2\right)^{-1} \left(\frac{1}{T} s_{v\omega} + \frac{1}{T} X_2' P_1 X_2 \hat{\beta}\right) \Rightarrow$$

$$\text{(3.33)} \quad \operatorname*{plim}_{T\to\infty} \hat{\beta}_S = \operatorname*{plim}_{T\to\infty} \hat{\beta}$$

und

$$\operatorname*{plim}_{T\to\infty} \hat{\sigma}_S^2 = \operatorname*{plim}_{T\to\infty} \hat{\sigma}^2 + \operatorname*{plim}_{T\to\infty} \frac{s_{\omega\omega}}{T-m-n} + \operatorname*{plim}_{T\to\infty} \frac{\hat{\beta}' X_2' P_1 X_2 \hat{\beta}}{T-m-n}$$

$$- \operatorname*{plim}_{T\to\infty} \left(\frac{s_{v\omega}}{T-m-n} + \frac{X_2' P_1 X_2 \hat{\beta}}{T-m-n}\right) \left(\frac{S_v}{T} + \frac{X_2' P_1 X_2}{T}\right)^{-1} \left(\frac{s_{v\omega}}{T} + \frac{X_2' P_1 X_2 \hat{\beta}}{T}\right)$$

woraus folgt:

$$\text{(3.34)} \quad \operatorname*{plim}_{T\to\infty} \hat{\sigma}_S^2 = \operatorname*{plim}_{T\to\infty} \hat{\sigma}^2$$

wobei im multiplen Modell $\hat{\beta}$ wie in (3.9) und $\hat{\sigma}$ wie in (2.23) definiert sind.

Nach wie vor ist also die Bedeutung der K-Q-Schätzungen aus Bayesscher Sicht im Falle großer Stichproben ungebrochen, denn da nur der zweite Faktor in (3.30) von Umfang und Ergebnis der Stichprobe abhängt und die Form einer multivariaten t-Dichtefunktion hat und der erste Faktor im Bereich $-\infty < \beta < +\infty$ endlich ist, konvergiert für $T\to\infty$ die a posteriori Wahrscheinlichkeitsverteilung von β gegen die Einpunktverteilung an der Stelle des Grenzwertes der K-Q-Schätzung von β.

3.3 A priori Information über Regressionskoeffizienten und Kovarianzmatrix der Fehler

Wenn wir schon a priori Information über die Kovarianzmatrix der Fehler als gegeben annehmen, so werden wir in der Regel auch a priori Kenntnisse über die Koeffizienten unterstellen dürfen - zumindest, wenn die a priori Informationen aus früheren empirischen Untersuchungen des gleichen Modells stammen.

Zunächst könnte man daran denken, als a priori Dichtefunktion z.B. für ß eine multiple t-Verteilung wie in (3.1) und für Σ eine inverse Wishart-Verteilung wie in (3.24) anzunehmen, und es ergäbe sich als a posteriori Randdichtefunktion von ß (gemäß (3.30)):

$$(3.35)\quad p_1(\beta|D,T_o,\hat{\beta}_o,\hat{\Sigma}_o) \propto \{1 + \frac{(\beta-\hat{\beta}_o)'\hat{\Sigma}_o^{-1}(\beta-\hat{\beta}_o)}{(T_o-m-n)}\}^{-\frac{T_o-m}{2}}$$

$$\{1 + \frac{(\beta-S_v^{-1}s_{v\omega})'S_v(\beta-S_v^{-1}s_{v\omega})}{s_{\omega\omega}-s_{v\omega}'S_v^{-1}s_{v\omega}}\}^{\frac{k-n}{2}}$$

$$\{1 + \frac{(\beta-\hat{\beta}_S)'(S_v+X_2'P_1X_2)(\beta-\hat{\beta})}{(T-m-n)\hat{\sigma}_S^2}\}^{-\frac{T+k-m-n}{2}}$$

wobei $\hat{\beta}_S$ und $\hat{\sigma}_S$ wie in (3.31) und (3.32) definiert sind; wenn man so will also eine verallgemeinerte "Dreifach-t-Verteilung". Dies erscheint aber gerade, wenn man die Möglichkeit offen halten möchte, daß die Information aus früheren Untersuchungen des gleichen Zusammenhangs stammt, nicht konsequent. Haben wir im Falle von a priori Information über ß allein die a priori Dichtefunktion aus derselben Familie gewählt zu der die a posteriori Dichte bei Unkenntnis über die Parameter gehört und im Falle von a priori Information über Σ allein auch auf die Familie der Dichtefunktionen zurückgegriffen, der

die a posteriori Verteilung von Σ bei gegebenen Koeffizienten und Realisationen der fehlerhaft gemessenen Variablen bei a priori Unkenntnis über Σ angehört, so wäre folgerichtig auch als a priori Dichte für die Regressionskoeffizienten und Σ zusammen eine Dichtefunktion aus derselben Klasse zu wählen, zu der die a posteriori Dichtefunktion für α, β und Σ bei a priori Unkenntnis gehört.

Beschränken wir uns auf den Fall der Unabhängigkeit zwischen V und ω, so wäre demnach als a priori Dichte entsprechend den Gleichungen (2.25) und (2.26) zu wählen:

$$(3.36)\quad p_o(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega}|y_o,X_{1o},X_{2o},T_o) \propto |\Sigma_v|^{-\frac{n+1}{2}} \sigma_{\omega\omega}^{-1}(\sigma_{\omega\omega}+\beta'\Sigma_v\beta)^{-\frac{T_o}{2}} e^{-\frac{(y_o - X_{1o}\alpha - X_{2o}\beta)'(y_o - X_{1o}\alpha - X_{2o}\beta)}{2(\sigma_{\omega\omega} + \beta'\Sigma_v\beta)}}.$$

In dieser Form haben wir eine a priori Dichte, die durch Informationen aus früheren Stichproben gewonnen sein kann. Dagegen bleibt offen wie beispielsweise im Falle der inversen Wishart-Verteilung als a priori Verteilung S gewonnen werden könnte.

Lassen wir es bezüglich der Elemente von Ξ bei diffuser Gleichverteilung bewenden, so erhalten wir a posteriori aus (3.36) mit der Likelihood (2.7) nach Integration über die Elemente der Matrix Ξ gemäß (2.25):

$$(3.37)\quad p_1(\alpha,\beta,\Sigma_v,\sigma_{\omega\omega}|D) \propto |\Sigma_v|^{-\frac{n+1}{2}} \sigma_{\omega\omega}^{-1} (\sigma_{\omega\omega}+\beta'\Sigma_v\beta)^{-\frac{T+T_o}{2}} e^{-\frac{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta) + (y_o-X_{1o}\alpha-X_{2o}\beta)'(y_o-X_{1o}\alpha-X_{2o}\beta)}{2(\sigma_{\omega\omega} + \beta'\Sigma_v\beta)}}$$

wobei D wieder die Daten und die a priori gegebenen Parameter abkürzend zusammenfassen soll. Substituieren wir Σ_v durch $\Sigma_v^{\times} = \sigma_{\omega\omega}^{-1}\Sigma_v$ mit der Jakobischen Determinante:

$$J = \sigma_{\omega\omega}^{-\frac{n(n+1)}{2}},$$

so erhalten wir:

$$(3.38)\quad p_1(\alpha,\beta,\Sigma_v^{\times},\sigma_{\omega\omega}|D) \propto |\Sigma_v^{\times}|^{-\frac{n+1}{2}} \sigma_{\omega\omega}^{-\frac{T+T_o}{2}-1} (1+\beta'\Sigma_v^{\times}\beta)^{-\frac{T+T_o}{2}} \cdot e^{-\frac{(y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)+(y_o-X_{1o}\alpha-X_{2o}\beta)'(y_o-X_{1o}\alpha-X_{2o}\beta)}{2\sigma_{\omega\omega}(1+\beta'\Sigma_v^{\times}\beta)}},$$

und die Integration über $\sigma_{\omega\omega}$ über den Bereich der positiven reellen Zahlen ergibt:

$$(3.39)\quad p_1(\alpha,\beta,\Sigma_v^{\times}|D) \propto |\Sigma_v^{\times}|^{-\frac{n+1}{2}} \left\{1+\frac{\begin{pmatrix}\alpha-\hat{\hat{\alpha}}\\ \beta-\hat{\hat{\beta}}\end{pmatrix}'(X_o'X_o+X'X)\begin{pmatrix}\alpha-\hat{\hat{\alpha}}\\ \beta-\hat{\hat{\beta}}\end{pmatrix}}{(T+T_o-m-n)\hat{\hat{\sigma}}^2}\right\}^{-\frac{T+T_o}{2}}$$

wobei: $\begin{pmatrix}\hat{\hat{\alpha}}\\ \hat{\hat{\beta}}\end{pmatrix} = (X_o'X_o+X'X)^{-1}(X_o'y_o+X'y)$ und

$$(T+T_o-m-m)\hat{\hat{\sigma}}^2 = y_o'y_o+y'y-\begin{pmatrix}\hat{\hat{\alpha}}\\ \hat{\hat{\beta}}\end{pmatrix}'(X_o'X_o+X'X)\begin{pmatrix}\hat{\hat{\alpha}}\\ \hat{\hat{\beta}}\end{pmatrix}$$

mit $X = (X_1,X_2)$ und $X_o = (X_{1o},X_{2o})$.

In bezug auf $\Sigma_v^{\times}$ enthält (3.39) also keine Information, was daran liegt, daß (3.36) keine Information über $\Sigma_v^{\times}$ enthält und auch aus der Stichprobe keine Information über $\Sigma_v^{\times}$ gewonnen werden kann, wie wir schon in Abschnitt 2.2 sahen. Letztlich können wir aus den Daten über $\Sigma_v^{\times}$ also nichts erfahren. Informationen über Σ sind allenfalls durch direkte Betrachtung der Meß- bzw Schätzmethoden und durch Überprüfungen der Messungen zu erreichen.

Hätten wir für $\Sigma_v^{\times}$ eine echte Dichtefunktion angenommen, so stünde diese an der Stelle von $|\Sigma_v^{\times}|^{-(n+1)/2}$ und als Randdichtefunktion von α und β bliebe der zweite Faktor in (3.39) übrig, eine (m+n)variate t-Verteilung, deren Mittelwert auch wieder nur bei kleinem Stichprobenumfang

von den K-Q-Schätzungen aus den Stichprobendaten für α und β abweicht. Dasselbe gilt für die Kovarianzmatrix. Faßt man die Daten und die a priori Information zusammen, also X_o und X zu $Z' := (X_o', X')$ sowie y_o und y zu $\overset{o}{y}' = (y_o', y')$, so ergeben sich formal sogar bei beliebigem Stichprobenumfang T Kleinstquadratschätzer:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (Z' Z)^{-1} Z' \overset{o}{y} \quad \text{und}$$

$$(T + T_o - m - n)\hat{\sigma}^2 = \overset{o}{y}' (I - Z(Z' Z)^{-1} Z') \overset{o}{y} .$$

Übrigens könnte man dieselbe a priori Dichte wie in (3.36) und ebenso die Verteilung (3.39) auch mit a priori gegebenen K-Q-Schätzungen $\hat{\gamma}_o := \begin{pmatrix} \hat{\alpha}_o \\ \hat{\beta}_o \end{pmatrix}$, $\hat{\Sigma}_{\hat{\gamma}o}$ und $\hat{\sigma}_o^2$ sowie mit T_o formulieren, sofern $X_o' X_o$ eine reguläre Matrix ist, denn aus $\hat{\Sigma}_{\hat{\gamma}o} = (X_o' X_o)^{-1} \hat{\sigma}_o^2$ folgt $X_o' X_o = \hat{\sigma}_o^2 \hat{\Sigma}_{\hat{\gamma}o}^{-1}$, aus

$\hat{\gamma}_o = (X_o' X_o)^{-1} X_o' y_o$ folgt $X_o' y_o = \hat{\sigma}_o^2 \hat{\Sigma}_{\hat{\gamma}o}^{-1} \hat{\gamma}$ und aus

$(T_o - m - n)\hat{\sigma}_o^2 = y_o' (I - X_o (X_o' X_o)^{-1} X_o') y_o$ folgt

$y_o' y_o = (T_o - m - n)\hat{\sigma}_o^2 + \hat{\gamma}' \hat{\Sigma}_{\hat{\gamma}o}^{-1} \hat{\gamma} \hat{\sigma}_o^2$ und nur diese Matrizen bzw. Vektoren aus Produktsummen benötigt man wirklich - sie sind zusammen mit den entsprechenden Stichprobenmomenten suffizient.

3.4 A priori Information über die fehlerhaft beobachteten erklärenden Variablen

Bisher haben wir stets die Elemente von Ξ als vollkommen unbekannt in dem Sinne betrachtet, daß wir über Ξ a priori eine diffuse Gleichverteilung angenommen haben. In allen im vorangegangenen untersuchten Fällen verschiedener a priori Verteilungen über die Regressionskoeffizienten und die Kovarianzmatrix der Fehler spielen in den a posteriori Randdichten der Regressionskoeffizienten die K-Q-Schätzungen eine dominierende Rolle insoweit als der Mittelwert der a posteriori Randverteilung entweder

gleich der K-Q-Schätzung der Koeffizienten der Regression von y auf X ist oder aber immerhin gegen den Wahrscheinlichkeitslimes der K-Q-Schätzung mit Wahrscheinlichkeit konvergiert.

Die bekannte Inkonsistenz der K-Q-Schätzungen im Falle von Fehlern in den Variablen beruht aber auf einer Annahme gerade über Ξ : Wie schon erwähnt soll nämlich:

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\,\Xi'\Xi = M_{\Xi} < \infty$$

sein, womit dann

$$\text{(3.40)}\quad \operatorname*{plim}_{T\to\infty} \hat{\beta} = (M_{\Xi} + \Sigma_v - M_{12}'M_{11}^{-1}M_{12})^{-1}\,(M_{\Xi} - M_{12}'M_{11}^{-1}M_{12})\,\beta + (M_{\Xi} + \Sigma_v - M_{12}'M_{11}^{-1}M_{12})^{-1}\sigma_{v\omega}$$

$$\neq \beta$$

wobei $\operatorname*{plim}_{T\to\infty} \frac{1}{T}\,X_1'X_1 = M_{11}$ und $\operatorname*{plim}_{T\to\infty} \frac{1}{T}\,X_1'\Xi = M_{12}$.

Ist a priori über Ξ nichts bekannt und eine diffuse Gleichverteilung über Ξ die adäquate Form der Berücksichtigung dieser Unkenntnis, so ist aber gerade diese Annahme über Ξ a priori nicht erfüllt und daher die K-Q-Schätzung als Mittelwert der a posteriori Verteilung akzeptabel.

Betrachten wir noch kurz das andere Extrem: Ist Ξ exakt bekannt, so entfällt das eigentliche Problem der Fehler in den Beobachtungen, denn wir können dann auf X_2 verzichten und allein die Beziehung

$$y = X_1\alpha + \Xi\beta + \omega$$

verwenden. Mit den Annahmen A1 und A2 erhalten wir bei diffuser a priori Verteilung hinsichtlich α, β und Σ für die Regressionskoeffizienten fast dieselben Verteilungen wie in (2.22) und (2.30). Der Unterschied liegt lediglich darin, daß gegenüber (2.22) und (2.30) die fehlerhaften Beobachtungen X_2 durch die bekannte Matrix Ξ ersetzt

werden müssen.[25] Wir können in diesem Fall natürlich Informationen über Σ gewinnen, worauf aber hier nicht näher eingegangen werden soll.

Die K-Q-Schätzungen mit Verwendung der Beobachtungen X_2 können nicht mehr als Mittelwert der a posteriori Verteilung auftreten, wenn wir a priori über Ξ eine Verteilung annehmen, die sicherstellt, daß plim $\frac{1}{T}\Xi'\Xi < \infty$ ist. In diesem Abschnitt möchte ich mich auf das einfache Modell beschränken, also:

$$y = \alpha + \beta\xi + \omega \tag{2.6}$$

$$x = \xi + v \tag{2.1}$$

wobei ξ, ω und v Realisationen unabhängiger Zufallsvariablen seien. ω sei normalverteilt mit Mittelwert O und Kovarianzmatrix $\sigma_\omega^2 I$ und v sei normalverteilt mit Mittelwert O und Kovarianzmatrix $\sigma_v^2 I$. Nehmen wir a priori an, daß:

$$p_o(\alpha,\beta,\sigma_v^2,\sigma_\omega^2,\xi|\mu,\sigma_\xi^2) \propto p_o(\alpha,\beta,\sigma_v^2,\sigma_\omega^2)\sigma_\xi^{-T} e^{-\frac{(\xi-\mu\iota)'(\xi-\mu\iota)}{2\sigma_\xi^2}} \tag{3.41}$$

so erhalten wir a posteriori:

$$p_1(\alpha,\beta,\sigma_v^2,\sigma_\omega^2,\xi|x,y,\mu,\sigma_\xi^2) \propto \frac{p_o(\alpha,\beta,\sigma_v^2,\sigma_\omega^2)}{(\sigma_\xi\sigma_v\sigma_\omega)^T} e^{-\frac{a}{2\sigma_\omega^2\sigma_v^2\sigma_\xi^2}} \tag{3.42}$$

wobei: $a = \sigma_v^2\sigma_\omega^2(\xi-\mu\iota)'(\xi-\mu\iota)+\sigma_\xi^2\sigma_\omega^2(x-\xi)'(x-\xi)+\sigma_\xi^2\sigma_v^2(y^x-\xi\beta)'(y^x-\xi\beta)$

mit $y^x = y - \alpha\iota$ und $\iota' = (1,1,\ldots,1)$.

Die Integration über ξ ergibt:

$$p_1(\alpha,\beta,\sigma_v^2,\sigma_\omega^2|x,y,\mu,\sigma_\xi^2) \propto \frac{p_o(\alpha,\beta,\sigma_v^2,\sigma_\omega^2)\, e^{-\frac{b}{2(\sigma_v^2\sigma_\omega^2+\sigma_\xi^2\sigma_\omega^2+\sigma_\xi^2\sigma_v^2\beta^2)}}}{(\sigma_v^2\sigma_\omega^2+\sigma_\xi^2\sigma_\omega^2+\sigma_\xi^2\sigma_v^2\beta^2)^{\frac{T}{2}}} \tag{3.43}$$

---

25) Zur Bayesschen Analyse des Regressionsmodells ohne Fehler in den Variablen siehe: A. Zellner und G.C. Tiao (1964)

wobei $b := \sigma_\xi^2 (y^\times - x\beta)'(y^\times - x\beta) + \sigma_v^2 (y^\times - \iota\mu_\beta)'(y^\times - \iota\mu_\beta) + \sigma_\omega^2 (x - \mu\iota)'(x - \mu\iota)$.

Weitere analytische Integration über σ_v^2 oder σ_ω^2 zur Bestimmung der Randverteilung von α und β scheint nicht durchführbar. Ist dagegen statt des absoluten Wertes der a priori Varianz von ξ das Verhältnis dieser Varianz zur Varianz der Fehler in den Beobachtungen von ξ, also $\sigma_\xi^2 / \sigma_v^2 =: \sigma_\xi^{\times 2}$ bekannt und wählen wir statt σ_v^2 den Parameter $\sigma_v^{\times 2} := \sigma_v^2 / \sigma_\omega^2$, so gilt:

$$(3.44)\quad p_1(\alpha,\beta,\sigma_v^{\times 2},\sigma_\omega^2 | x,y,\mu,\sigma_\xi^{\times 2}) \propto \frac{p_o(\alpha,\beta,\sigma_v^{\times 2},\sigma_\omega^2)}{\sigma_\omega^{2T}\sigma_v^{\times T}(1+\sigma_\xi^{\times 2}+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\beta^2)^{\frac{T}{2}}} e^{-\frac{b}{2\sigma_\omega^2\sigma_v^{\times 2}(1+\sigma_\xi^{\times 2}+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\beta^2)}}$$

wobei $b^\times := \sigma_\xi^{\times 2}\sigma_v^{\times 2}(y^\times - x\beta)'(y^\times - x\beta) + \sigma_v^{\times 2}(y^\times - \iota\mu_\beta)'(y^\times - \iota\mu_\beta) + (x - \iota\mu)'(x - \iota\mu)$.

Mit

$$(3.45)\quad p_o(\alpha,\beta,\sigma_v^{\times 2},\sigma_\omega^2) \propto \frac{1}{\sigma_\omega^2}\ p_o(\alpha,\beta,\sigma_v^{\times 2})$$

liefert die Integration über σ_ω^2 (gemäß Anhang 2.2):

$$(3.46)\quad p_1(\alpha,\beta,\sigma_v^{\times 2} | x,y,\mu,\sigma_\xi^{\times 2}) \propto p_o(\alpha,\beta,\sigma_v^{\times 2})\sigma_v^{\times T}(1+\sigma_\xi^{\times 2}+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\beta^2)^{\frac{T}{2}} b^{\times -T}$$

Leider ist mit vager a priori Dichte über $\sigma_v^{\times 2}$ auch die a posteriori Verteilung in bezug auf $\sigma_v^{\times 2}$ nicht echt. Integrieren wir daher nun noch das weniger interessierende konstante Glied α unter der Annahme

$$(3.47)\quad p_o(\alpha,\beta,\sigma_v^{\times 2}) \propto p_o(\beta,\sigma_v^{\times 2})$$

heraus. Zu diesem Zweck formen wir $b^\times$ wie folgt um:

$$b^{\times} = T\sigma_v^{\times 2}(\sigma_\xi^{\times 2}+1)\{\alpha - \frac{\sigma_\xi^{\times 2}\sum_{t=1}^{T}(y_t-\beta x_t)+\sum_{t=1}^{T}(y_t-\beta\mu)}{T(\sigma_\xi^{\times 2}+1)}\}^2$$

$$+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\sum_{t=1}^{T}(y_t-\beta x_t)^2+\sigma_v^{\times 2}\sum_{t=1}^{T}(y_t-\beta\mu)^2+\sum_{t=1}^{T}(x_t-\mu)^2$$

$$-\frac{\sigma_v^{\times 2}}{T(\sigma_\xi^{\times 2}+1)}\{\sigma_\xi^{\times 4}[\sum_{t=1}^{T}(y_t-\beta x_t)]^2+2\sigma_\xi^{\times 2}\sum_{t=1}^{T}(y_t-\beta x_t)\sum_{t=1}^{T}(y_t-\beta\mu)$$

$$+[\sum_{t=1}^{T}(y_t-\beta\mu)]^2\} \ .$$

Unter Ausnutzung der Eigenschaften der t-Verteilung (α ist bei gegebenem $\sigma_v^{\times 2}$, β sowie $\sigma_\xi^{\times 2}$ und μ a posteriori t-verteilt) folgt:

$$(3.48)\quad p_1(\beta,\sigma_v^{\times 2}\,|\,x,y,\mu,\sigma_\xi^{\times 2}) \propto p_0(\beta,\sigma_v^{\times 2})\,\sigma_v^{\times T-1}(1+\sigma_\xi^{\times 2}+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\beta^2)^{-\frac{T}{2}}$$

$$\Big[\{T(\sigma_\xi^{\times 2}+1)\sigma_v^{\times 2}(\sigma_\xi^{\times 2}\sum_{t=1}^{T}x_t^2+T\mu^2)$$

$$-\sigma_v^{\times 2}(\sigma_\xi^{\times 2}\sum_{t=1}^{T}x_t+T\mu)^2\}(\beta-\tilde{\beta})^2$$

$$+(\sigma_\xi^{\times 2}+1)^2\sigma_v^{\times 2}[T\sum_{t=1}^{T}y_t^2-(\sum_{t=1}^{T}y_t)^2]$$

$$+(\sigma_\xi^{\times 2}+1)T\sum_{t=1}^{T}(x_t-\mu)^2$$

$$-\sigma_v^{\times 2}\{T(\sigma_\xi^{\times 2}+1)(\sigma_\xi^{\times 2}\sum_{t=1}^{T}x_t^2+T\mu^2)$$

$$-(\sigma_\xi^{\times 2}\sum_{t=1}^{T}x_t+T\mu)^2\}\tilde{\beta}^2\Big]^{-T+\frac{1}{2}}$$

wobei

$$(3.49)\qquad \tilde{\beta} = \frac{T\sum_{t=1}^{T}x_ty_t-\sum_{t=1}^{T}x_t\sum_{t=1}^{T}x_t}{\{T\sum_{t=1}^{T}x_t^2-(\sum_{t=1}^{T}x_t)^2\}+\frac{(\sum x_t-T\mu)^2}{\sigma_\xi^{\times 2}+1}} \ .$$

Es gilt also:

$$(3.50)\quad p_1(\beta|\sigma_v^{\times 2},\sigma_\xi^{\times 2},\mu,x,y) \propto p_o(\beta|\sigma_v^{\times 2})(1+\sigma_\xi^{\times 2}+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\beta^2)^{\frac{T}{2}} \{1+\frac{(\beta-\tilde{\beta})^2}{(2T-2)\hat{\sigma}_{\tilde{\beta}}^2}\}^{-\frac{2T-1}{2}}$$

wobei

$$(3.51)\quad (2T-2)\hat{\sigma}_{\tilde{\beta}}^2 = \frac{(\sigma_\xi^{\times 2}+1)\{(\sigma_\xi^{\times 2}+1)\sigma_v^{\times 2}\{T\sum_{t=1}^{T} y_t^2-(\sum_{t=1}^{T} y_t)^2\}+T\sum_{t=1}^{T}(x_t-\mu)^2\}}{\sigma_\xi^{\times 2}\sigma_v^{\times 2}\{(\sigma_\xi^{\times 2}+1)\{T\sum_{t=1}^{T} x_t^2-(\sum_{t=1}^{T} x_t)^2\}+(\sum_{t=1}^{T} x_t-T\mu)^2\}} - \tilde{\beta}^2.$$

Zu bemerken ist, daß $\tilde{\beta}$ keine konsistente Schätzung für β ist, selbst wenn $\operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \xi_t = \mu$ ist, sondern es gilt:

$$\operatorname*{plim}_{T\to\infty} \tilde{\beta} = \beta \frac{\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \xi_t^2 - \{E(\xi)\}^2}{\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \xi_t^2 - \{E(\xi)\}^2 + \sigma_v^2 + \frac{\{E(\xi)-\mu\}^2}{\sigma_\xi^{\times 2}+1}}$$

woraus für $E(\xi) = \mu$ folgt:

$$(3.52)\quad \operatorname*{plim}_{T\to\infty} \tilde{\beta} = \operatorname*{plim}_{T\to\infty} \hat{\beta}\ .$$

In der Tat ist die einfache Kleinstquadratschätzung $\hat{\beta}$ der in (3.49) definierten noch vorzuziehen, denn es ist $|\tilde{\beta}| \leq |\hat{\beta}|$. Der Mittelwert und der Modalwert der a posteriori Verteilung (3.50) werden aber durch den Faktor $(1+\sigma_\xi^{\times 2}+\sigma_\xi^{\times 2}\sigma_v^{\times 2}\beta^2)^{T/2}$ zu - dem Betrage nach - größeren Werten als $\tilde{\beta}$ verschoben; und zwar nicht nur bei kleinen Stichproben. Auch unter der Annahme von a priori Wissen über die Kovarianzmatrix der Fehler tritt, wie wir gesehen haben, ein Faktor auf, der bei kleiner Stichprobe den Modal- und den Mittelwert der a posteriori Dichte von β von dem Maximum des Faktors, der dem Kern einer t-Dichtefunktion gleich ist, weg verschiebt(vgl. zum

Beispiel die Beziehung (3.30^{x})). Im Gegensatz zu dort verstärkt sich aber hier die "Korrektur" mit wachsendem Stichprobenumfang - der Exponent des Korrekturfaktors ist gleich T/2 - und bleibt damit auch bei großen Stichproben bedeutsam.

4 Übersicht über die gewonnenen Ergebnisse

Es wurde in der vorliegenden Arbeit der Versuch unternommen, das Schätzproblem bei normalverteilten Fehlern in den Variablen durch Hinzunahme verschiedenartiger a priori Information auf dem Weg über das Bayessche Theorem einer Lösung näher zu bringen.

Das Hauptgewicht der Untersuchungen lag dabei auf der Einbeziehung unvollständigen a priori Wissens - in diesem Zusammenhang zu verstehen als echte Information in Gestalt einer a priori Dichtefunktion. Demgemäß wurde ausgehend von dem Fall völlig vager a priori Kenntnis, die ihren Ausdruck in diffusen a priori Verteilungen fand, echte a priori Information nacheinander über die Regressionskoeffizienten selbst, über die Kovarianzmatrix der Fehler und über die sogenannten Inzidenzparameter, also die fehlerhaft beobachteten Variablen, eingeführt und die a posteriori Randverteilung des Koeffizientenvektors - wenn möglich - berechnet.

Es zeigte sich, daß häufig die analytische Bestimmung dieser a posteriori Randverteilung der Koeffizienten auf enorme Schwierigkeiten stößt, und es wurde häufig notwendig, eine Änderung der Parametrisierung vorzunehmen, was letztlich auf eine Veränderung der Modellannahmen hinausläuft, wenn eben die Berücksichtigung der entsprechenden Funktionaldeterminanten dadurch umgangen wurde, daß - gewissermaßen rückwirkend - die a priori Verteilung in diesen anderen Parametern definiert wurde.

In überraschend vielen Fällen spielen die Kleinstquadratschätzungen entscheidende Rollen, insofern als die Mittelwerte der a posteriori Randdichten entweder gleich den K-Q-Schätzungen sind oder aber mit wachsendem Stichprobenumfang der Wahrscheinlichkeit nach gegen sie konvergieren. Lediglich im Falle exakter Kenntnis

der Fehlerkovarianzmatrix (vgl.(3.7)) und bei echter a priori Information über die Inzidenzparameter (vgl. (3.50)) ist auch bei großen Stichproben der a posteriori Mittelwert der Randverteilung der Koeffizienten fehlerhaft gemessener Variablen nicht gleich der K-Q-Schätzung.

Die folgende Übersicht möge die gewonnenen Ergebnisse im Zusammenhang mit der jeweiligen a priori Wahrscheinlichkeitsverteilung zusammenfassen.

Übersicht über die unter verschiedenen a priori Annahmen gewonnenen Ergebnisse

a priori			a posteriori		
1	2	3	4	5	6
unbekannte Parameter	a priori Verteilung in bezug auf	a priori geg. Parameter	unbekannte Parameter	a posteriori Verteilung bzw. Dichtefunktion[1] der Unbekannten in Spalte 4	Parameter der a post.V.[2]
$\alpha,\beta,\Xi,$ $\Sigma^{\times},\sigma_{\omega\omega}$	Ξ und $\log\sigma_{\omega\omega}$:diffus α,β und $\Sigma^{\times}$: nicht spezifiziert		$\alpha,\beta,\Sigma^{\times}$	$p_o(\alpha,\beta,\Sigma^{\times})\left\{1+\frac{1}{(T-m-n)\hat{\sigma}^2}\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}' X'X \begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}\right\}^{-\frac{T}{2}}$	$\hat{\alpha},\hat{\beta},$ $\hat{\sigma}^2$
α,β,Ξ $\Sigma_v^{\times},\sigma_{\omega\omega}$	α und β: multivariat t-verteilt mit T_o-m-n Freiheitsgraden Ξ und $\log\sigma_{\omega\omega}$: diffus $\Sigma_v^{\times}$: nicht spezifiziert	$\hat{\alpha}_o,\hat{\beta}_o,$ $\hat{\Sigma}_o,T_o$ $\sigma_{v\omega}=0$	α,β	$\left\{1+\frac{\begin{pmatrix}\alpha-\hat{\alpha}_o\\ \beta-\hat{\beta}_o\end{pmatrix}'\hat{\Sigma}_o^{-1}\begin{pmatrix}\alpha-\hat{\alpha}_o\\ \beta-\hat{\beta}_o\end{pmatrix}}{(T_o-m-n)}\right\}^{-\frac{T_o}{2}}\left\{1+\frac{\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}' X'X\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}}{(T-m-n)\hat{\sigma}^2}\right\}^{-\frac{T}{2}}$	$\hat{\alpha},\hat{\beta},$ $\hat{\sigma}^2$
α,β,Ξ	α, β und Ξ : diffus	Σ	β	$\left\{\begin{pmatrix}\beta\\ -1\end{pmatrix}'\Sigma\begin{pmatrix}\beta\\ -1\end{pmatrix}\right\}^{-\frac{T-m}{2}}\exp\left\{-\frac{(\beta-\hat{\beta})'X_2'P_1X_2(\beta-\hat{\beta})+y'P_1P_{21}^{-1}P_1y}{2\begin{pmatrix}\beta\\ -1\end{pmatrix}'\Sigma\begin{pmatrix}\beta\\ -1\end{pmatrix}}\right\}$	$\hat{\beta}$
α,β,Ξ $\Sigma_v^{\times},$	$\sigma_{\omega\omega}$: invers χ^2 mit k Freiheitsgraden Ξ,α und β : diffus $\Sigma_v^{\times}$: nicht spezifiziert	s , k $\sigma_{v\omega}=0$	α , β $\Sigma_v^{\times}$	$p_o(\Sigma_v^{\times})(1+\beta'\Sigma_v^{\times}\beta)^{\frac{k}{2}}\left\{(T-m-n)\hat{\sigma}^2+(1+\beta'\Sigma_v^{\times}\beta)\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}'\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}\right\}^{-\frac{T+k}{2}}$	$\hat{\alpha}$, $\hat{\beta}$ $\hat{\sigma}^2$
α,β,Ξ $\sigma_{\omega\omega}$	$\sigma_{\omega\omega}$: invers χ^2 mit k Freiheitsgraden Ξ,α und β: diffus	s,k, $\Sigma_v^{\times}$ $\sigma_{v\omega}=0$	β	$(1+\beta'\Sigma_v^{\times}\beta)^{\frac{k}{2}}\left\{1+\frac{(\beta-\hat{\beta}_s)'(s\Sigma_v^{\times}+X_2'P_1X_2)(\beta-\hat{\beta}_s)}{s+(T-m-n)\hat{\sigma}^2+\hat{\beta}'X_2'P_1M^{-1}P_1X_2\hat{\beta}}\right\}^{-\frac{T+k-m}{2}}$ wobei $M:=I-X_2(s\Sigma_v^{\times}+X_2'P_1X_2)^{-1}X_2'$	$\hat{\beta}_s$, $\hat{\beta}$, $\hat{\sigma}^2$

1) bis auf die Normierungskonstante

2) außer den a priori gegebenen Parametern, den Matrizen $X_1'X_1$, $X_1'X_2$ und $X_2'X_2$ sowie den Konstanten T, m und n

1 unbekannte Parameter	2 a priori Verteilung in bezug auf	3 a priori geg. Parameter	4 unbekannte Parameter	5 a posteriori Verteilung bzw. Dichtefunktion der Unbekannten in Spalte 4	6 Parameter der a post.V.
$\alpha, \beta,$ Ξ, Σ	Σ: invers Wishart Ξ: diffus α und β nicht spezifiziert	S und k	α , β	$p_o(\alpha,\beta)\{(\underset{-1}{\beta})' S(\underset{-1}{\beta})\}^{\frac{k-n}{2}} \left\{(\underset{-1}{\beta})' S(\underset{-1}{\beta}) + (T-m-n)\hat{\sigma}^2 + \begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}' \begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix} \begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}\right\}^{-\frac{T+k-n}{2}}$	$\hat{\alpha}, \hat{\beta}$ $\hat{\sigma}^2$
$\alpha, \beta,$ Ξ, Σ	Σ: invers Wishart Ξ, α und β: diffus	k $S=\begin{pmatrix}S_v & s_{v\omega}\\ s_{v\omega}' & s_{\omega\omega}\end{pmatrix}$	β	$\left(1+\frac{(\beta-S_v^{-1}s_{v\omega})' S_v (\beta-S_v^{-1}s_{v\omega})}{s_{\omega\omega} - s_{v\omega}' S_v^{-1} s_{v\omega}}\right)^{\frac{k-n}{2}} \times \left(1+\frac{(\beta-\hat{\beta}_S)'(S_v+X_2'P_1X_2)(\beta-\hat{\beta}_S)}{(T-m-n)\hat{\sigma}_S^2}\right)^{-\frac{T+k-n-m}{2}}$	$\hat{\beta}_S, \hat{\sigma}_S^2$
$\alpha, \beta,$ Ξ, Σ	Σ: invers Wishart β: multivariat t-verteilt mit T_o-m-n Freiheitsgraden Ξ und α: diffus	S, k $T_o, \hat{\beta}_o,$ $\hat{\Sigma}_{\hat{\beta}_o}$	β	$\left(1+\frac{(\beta-\hat{\beta}_o)'\hat{\Sigma}_{\hat{\beta}o}^{-1}(\beta-\hat{\beta}_o)}{T_o-m-n}\right)^{-\frac{T_o-m}{2}} \times \left(1+\frac{(\beta-S_v^{-1}s_{v\omega})' S_v (\beta-S_v^{-1}s_{v\omega})}{s_{\omega\omega} - s_{v\omega}' S_v^{-1} s_{v\omega}}\right)^{\frac{k-n}{2}} \times \left(1+\frac{(\beta-\hat{\beta}_S)'(S_v+X_2'P_1X_2)(\beta-\hat{\beta}_S)}{(T-m-n)\hat{\sigma}_S^2}\right)^{-\frac{T+k-n-m}{2}}$	$\hat{\beta}_S, \hat{\sigma}_S^2$
$\alpha, \beta,$ $\Xi, \Sigma_v^{\times},$ $\sigma_{\omega\omega}$	$\alpha, \beta, \Sigma_v^{\times}$ und $\sigma_{\omega\omega}$: entsprechend der a post. Verteilung bei a priori Unkenntnis Ξ: diffus	y_o, X_{1o} X_{2o}, T_o	α , β $\Sigma_v^{\times}$	$\lvert\Sigma_v^{\times}\rvert^{-\frac{n+1}{2}} \left\{1+\frac{\begin{pmatrix}\alpha-\hat{\hat{\alpha}}\\ \beta-\hat{\hat{\beta}}\end{pmatrix}' \begin{pmatrix}X_{1o}'X_{1o}+X_1'X_1 & X_{1o}'X_{2o}+X_1'X_2\\ X_{2o}'X_{1o}+X_2'X_1 & X_{2o}'X_{2o}+X_2'X_2\end{pmatrix} \begin{pmatrix}\alpha-\hat{\hat{\alpha}}\\ \beta-\hat{\hat{\beta}}\end{pmatrix}}{(T_o+T-m-n)\hat{\hat{\sigma}}^2}\right\}^{-\frac{T_o+T}{2}}$	$\hat{\hat{\sigma}}^2,$ $\hat{\hat{\alpha}},$ $\hat{\hat{\beta}},$

0	1	2	3	4	5	6
Formel Nr.	unbekannte Parameter	a priori Verteilung in bezug auf	a priori geg. Parameter	unbekannte Parameter	a posteriori Verteilung bzw. Dichtefunktion der Unbekannten in Spalte 4	Parameter der a post.V.
	$\alpha, \beta,$ $\sigma_{\omega\omega}$	$\log\sigma_{\omega\omega}$: diffus α und β nicht spezifiziert	Ξ	α , β	$p_o(\alpha,\beta)\left\{1+\frac{1}{(T-m-n)\hat{\sigma}^2_\Xi}\begin{pmatrix}\alpha-\hat{\alpha}_\Xi\\ \beta-\hat{\beta}_\Xi\end{pmatrix}'\begin{pmatrix}X_1'X_1 & X_1'\Xi\\ \Xi'X_1 & \Xi'\Xi\end{pmatrix}\begin{pmatrix}\alpha-\hat{\alpha}_\Xi\\ \beta-\hat{\beta}_\Xi\end{pmatrix}\right\}^{-\frac{T}{2}}$	$\hat{\alpha}_\Xi$, $\hat{\beta}_\Xi$ $\hat{\sigma}^2_\Xi$
3.50	$\alpha, \beta,$ $\xi,$ σ^2_ω	ξ: normal(μ,σ^2_ξ) α und $\log\sigma^2_\omega$: diffus β: nicht spezifiziert	μ , $\sigma^{\times 2}_\xi, \sigma^{\times 2}_v$	β	$p_o(\beta\mid\sigma^{\times 2}_v)\,(1+\sigma^{\times 2}_\xi+\sigma^{\times 2}_\xi\sigma^{\times 2}_v\beta^2)^{-\frac{T}{2}} \times\left\{1+\frac{(\beta-\tilde{\beta})^2}{(2T-2)\tilde{\sigma}^2_{\tilde{\beta}}}\right\}^{-\frac{2T-1}{2}}$	$\tilde{\beta}, \tilde{\sigma}^2_{\tilde{\beta}}$

Definitionen der Parameter der a posteriori Dichte
(aus Spalte 6 der Übersicht)

Definitionsgleichung	Bemerkungen
$\hat{\alpha} = (X_1'P_2X_1)^{-1}X_1'P_2y$ mit $P_2=I-X_2(X_2'X_2)^{-1}X_2'$	Kleinstquadrat-schätzungen der Koeffizienten der Regression $y = X_1\alpha + X_2\beta + u$
$\hat{\beta} = (X_2'P_1X_2)^{-1}X_2'P_1y$ mit $P_1=I-X_1(X_1'X_1)^{-1}X_1'$	
$\hat{\sigma}^2=\frac{1}{T-m-n}\, y'\{ I-(X_1,X_2)\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}^{-1}(X_1,X_2)'\}y$	K-Q-Schätzung der Varianz von u
$\hat{\beta}_s=(s\Sigma_v^{\times}+X_2'P_1X_2)^{-1}X_2'P_1X_2\hat{\beta}$	$\underset{T\to\infty}{\text{plim}}\hat{\beta}_s=\underset{T\to\infty}{\text{plim}}\hat{\beta}$
$\hat{\beta}_S=(S_v+X_2'P_1X_2)^{-1}(s_{v\omega}+X_2'P_1X_2\hat{\beta})$	$\underset{T\to\infty}{\text{plim}}\hat{\beta}_S=\underset{T\to\infty}{\text{plim}}\hat{\beta}$
$\hat{\sigma}_S^2=\frac{1}{T-m-n}\{(T-m-n)\hat{\sigma}^2+s_{\omega\omega}+\hat{\beta}'X_2'P_1X_2\hat{\beta}$ $-(s_{v\omega}+X_2'P_1X_2\hat{\beta})'(S_v+X_2'P_1X_2)^{-1}(s_{v\omega}+X_2'P_1X_2\hat{\beta})\}$	$\underset{T\to\infty}{\text{plim}}\hat{\sigma}_S^2=\underset{T\to\infty}{\text{plim}}\hat{\sigma}^2$
$\begin{pmatrix}\hat{\hat{\alpha}}\\ \hat{\hat{\beta}}\end{pmatrix}= (X_o'X_o+X'X)^{-1}(X_o'y_o+X'y_1)$ mit $X_o=(X_{1o},X_{2o})$ und $X=(X_1,X_2)$	K-Q-Schätzungen der Koeffizienten und der Varianz der Störvariablen der Regression $\begin{pmatrix}y_o\\ y\end{pmatrix}=\begin{pmatrix}X_{1o}\\ X_1\end{pmatrix}\alpha+\begin{pmatrix}X_{2o}\\ X_2\end{pmatrix}\beta+\begin{pmatrix}u_o\\ u\end{pmatrix}$
$\hat{\hat{\sigma}}^2=\frac{1}{T_o+T-m-n}\, y_o'y_o+y'y-\begin{pmatrix}\hat{\hat{\alpha}}\\ \hat{\hat{\beta}}\end{pmatrix}'(X'X+X'X)\begin{pmatrix}\hat{\hat{\alpha}}\\ \hat{\hat{\beta}}\end{pmatrix}$	
$\hat{\alpha}_\Xi= (X_1'P_\Xi X_1)^{-1}X_1'P_\Xi y$ mit $P_\Xi=I-\Xi(\Xi'\Xi)^{-1}\Xi'$	K-Q-Schätzungen der Koeffizienten und der Varianz der Störvariablen der Regression $y = X_1\alpha + \Xi\beta + \omega$ bei gegebenem Ξ
$\hat{\beta} = (\Xi'P_1\Xi)^{-1}\Xi'P_1y$	
$\hat{\sigma}^2=\frac{1}{T-m-n}\, y'\{I-(X_1,\Xi)\begin{pmatrix}X_1'X_1 & X_1'\Xi\\ \Xi'X_1 & \Xi'\Xi\end{pmatrix}^{-1}(X_1,\Xi)'\}y$	
$\tilde{\beta} = \dfrac{T\,\Sigma x_t y_t - \Sigma x_t \Sigma y_t}{T\Sigma x_t^2 - (\Sigma x_t)^2 + \dfrac{(\Sigma x_t - T\mu)^2}{\sigma_\xi^{\times 2}+1}}$	$\underset{T\to\infty}{\text{plim}}\tilde{\beta} = \underset{T\to\infty}{\text{plim}}\hat{\beta}$
$\hat{\sigma}_{\tilde{\beta}}^2=\frac{1}{T-2}\, \dfrac{(\sigma_\xi^{\times 2}+1)\{(\sigma_\xi^{\times 2}+1)\sigma_v^{\times 2}[T\Sigma y_t^2-(\Sigma y_t)^2]+T\Sigma(x_t-\mu)^2\}}{\sigma_\xi^{\times 2}\sigma_v^{\times 2}\{(\sigma_\xi^{\times 2}+1)[T\Sigma x_t^2-(\Sigma x_t)^2]+(\Sigma x_t-T\mu)^2\}} - \tilde{\beta}^2$	

Anhang 1: Zur Integration von (2.9) und (2.24) über Ξ

Satz 1 a) Ist $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, wobei A_{11} und A_{22} quadratisch seien und ist A_{11} regulär, so gilt:

$$|A| = |A_{11}||A_{22} - A_{21} A_{11}^{-1} A_{12}| \ .$$

b) Ist weiter A_{22} regulär, so gilt:

$$|A| = |A_{22}||A_{11} - A_{12} A_{22}^{-1} A_{21}| \ .$$

Beweis:
$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} I & -A_{11}^{-1}A_{12} \\ O & I \end{vmatrix} = \begin{vmatrix} A_{11} & O \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{vmatrix}$$
$\Rightarrow$ Behauptung.

Folgerung F1: Ist Σ eine reguläre $n \times n$ Matrix und B eine $n \times m$ Matrix, so gilt:

$$|\Sigma + BB'| = |\Sigma| \times |I + B'\Sigma^{-1}B| \ .$$

Beweis: Nach 1b) gilt mit $A_{11} = \Sigma$, $A_{12} = -B$, $A_{21} = B'$ und $A_{22} = I$: $\begin{vmatrix} \Sigma & -B \\ B' & I \end{vmatrix} = |\Sigma + BB'|$ und nach 1a) gilt:

$$\begin{vmatrix} \Sigma & -B \\ B & I \end{vmatrix} = |\Sigma||I - B'\Sigma^{-1}B|$$

$\Rightarrow$ Behauptung.

Folgerung F2: Ist Σ eine reguläre $(n+1) \times (n+1)$ Matrix und β ein n-Spaltenvektor, so gilt:

$$|(I_n, \beta)\Sigma^{-1}(I_n, \beta)'| = |\Sigma|^{-1} \begin{pmatrix} \beta \\ -1 \end{pmatrix}' \Sigma \begin{pmatrix} \beta \\ -1 \end{pmatrix} .$$

Beweis: Teilen wir Σ^{-1} wie folgt auf: $\Sigma^{-1} = \begin{pmatrix} \Sigma^{v} & \sigma^{v\omega} \\ \sigma^{v\omega}, & \sigma^{\omega\omega} \end{pmatrix}$, wobei Σ^{v} eine $m \times n$ Matrix, $\sigma^{v\omega}$ ein n-Spaltenvektor und $\sigma^{\omega\omega}$ ein Skalar sei, so gilt nach 1b):

$$\begin{vmatrix} \Sigma^{v} & \sigma^{v\omega} & \beta \\ \sigma^{v\omega\prime} & \sigma^{\omega\omega} & -1 \\ -\beta' & 1 & 0 \end{vmatrix} = \begin{vmatrix} \sigma^{\omega\omega} & -1 \\ 1 & 0 \end{vmatrix} \left| \Sigma^{v} - (\sigma^{v\omega}\beta)\begin{pmatrix} \sigma^{\omega\omega} & -1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} \sigma^{v\omega\prime} \\ -\beta' \end{pmatrix} \right|$$

$$= \left| (I_n, \beta)\Sigma^{-1}(I_n, \beta)' \right| .$$

Nach 1a) gilt:

$$\begin{vmatrix} \Sigma^{v} & \sigma^{v\omega} & \beta \\ \sigma^{v\omega\prime} & \sigma^{\omega\omega} & -1 \\ -\beta' & 1 & 0 \end{vmatrix} = \begin{vmatrix} \Sigma^{v} & \sigma^{v\omega} \\ \sigma^{v\omega} & \sigma^{\omega\omega} \end{vmatrix} \left| 0 - (-\beta', 1)\Sigma\begin{pmatrix} \beta \\ -1 \end{pmatrix} \right|$$

$$= \left|\Sigma^{-1}\right| \begin{pmatrix} \beta \\ -1 \end{pmatrix}'\Sigma\begin{pmatrix} \beta \\ -1 \end{pmatrix}$$

$\Rightarrow$ Behauptung.

Satz 2: Ist A eine reguläre n×n Matrix und sind b und c Spaltenvektoren mit jeweils n Komponenten, so gilt:

$$(A + bc')^{-1} = \left(A^{-1} - \frac{A^{-1}bc'A^{-1}}{1+c'A^{-1}b} \right) .^{1)}$$

Beweis:

$$(A+bc')\left(A^{-1} - \frac{A^{-1}bc'A^{-1}}{1+c'A^{-1}b} \right)$$

$$= I+bc'A^{-1} - \frac{bc'A^{-1}+bc'A^{-1}bc'A^{-1}}{1 + c'A^{-1}b}$$

$$= I + \frac{bc'A^{-1}c'A^{-1}b - bc'A^{-1}bc'A^{-1}}{1 + c'A^{-1}b}$$

$$= I$$

denn $c'A^{-1}b$ ist ein Skalar.

---

1) vgl. C.R. Rao (1965) S. 29 (Problem 2.8).

Satz 3: Ist $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ wobei A_{11} und A_{22} quadratisch und regulär sein sollen, so gilt für

$$A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}:$$

$$A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1},$$

$$A^{12} = - A^{11}A_{12}A_{22}^{-1},$$

$$A^{21} = - A_{22}^{-1}A_{21}A^{11} \quad \text{und}$$

$$A^{22} = A_{22}^{-1} - A_{22}^{-1}A_{21}A^{12}.$$

Beweis: A^{11} existiert, wie man durch Anwendung des Satzes 1b) sehen kann, und durch Ausmultiplizieren sieht man leicht, daß $A \cdot A^{-1} = I$ ist.

Satz 4: Ist Σ eine reguläre, symmetrische $(n+1)\times(n+1)$ Matrix und ist β ein Vektor mit n Komponenten, so gilt:

$$\left[(I_n,\beta)\Sigma^{-1}(I_n,\beta)'\right]^{-1} = \left[(I_n,\beta)(I_n,\beta)'\right]^{-1}(I_n,\beta)$$

$$\times \left(\Sigma - \frac{\Sigma\binom{\beta}{-1}\binom{\beta}{-1}'\Sigma}{\binom{\beta}{-1}'\Sigma\binom{\beta}{-1}}\right)(I_n,\beta)'\left[(I_n,\beta)(I_n,\beta)'\right]^{-1}$$

wobei I_n die $n\times n$ Einheitsmatrix ist.

Beweis: Unterteilen wir $\Sigma = \begin{pmatrix} \Sigma_v & \sigma_{v\omega} \\ \sigma_{v\omega}' & \sigma_{\omega\omega} \end{pmatrix}$ und $\Sigma^{-1} = \begin{pmatrix} \Sigma^v & \sigma^{v\omega} \\ \sigma^{v\omega}, & \sigma^{\omega\omega} \end{pmatrix}$

wobei $\sigma_{\omega\omega}$ und $\sigma^{\omega\omega}$ Skalare und Σ_v sowie Σ^v $n\times n$ Matrizen seien, so gilt nach Satz 3:

$$\Sigma^v = \left(\Sigma_v - \frac{\sigma_{v\omega}\,\sigma_{v\omega}'}{\sigma_{\omega\omega}}\right)^{-1} \quad <=>$$

$$\Sigma^v = \sigma_{\omega\omega}\{\sigma_{\omega\omega}\,\Sigma_v + (-\sigma_{v\omega})\sigma_{v\omega}'\}^{-1}$$

und nach Satz 2 gilt weiter:

$$\Sigma^{v} = \Sigma_{v}^{-1} + \frac{\Sigma_{v}^{-1}\sigma_{v\omega}\sigma_{v\omega}'\Sigma_{v}^{-1}}{\sigma_{\omega\omega} - \sigma_{v\omega}'\Sigma_{v}^{-1}\sigma_{v\omega}}$$

und weiter nach Satz 3:

$$\sigma^{v\omega} = -\frac{\Sigma_{v}^{-1}\sigma_{v\omega}}{\sigma_{\omega\omega} - \sigma_{v\omega}'\Sigma_{v}^{-1}\sigma_{v\omega}}$$

sowie:

$$\sigma^{\omega\omega} = \frac{1}{\sigma_{\omega\omega} - \sigma_{v\omega}'\Sigma_{v}^{-1}\sigma_{v\omega}} .$$

Damit ist:

$$(I_n,\beta)\Sigma^{-1}(I_n,\beta)' = \frac{1}{\sigma_{\omega\omega}-\sigma_{v\omega}'\Sigma_{v}^{-1}\sigma_{v\omega}} \{\sigma_{\omega\omega}\Sigma_{v}^{-1}-\sigma_{v\omega}'\Sigma_{v}^{-1}\sigma_{v\omega}\Sigma_{v}^{-1}$$

$$+\Sigma_{v}^{-1}\sigma_{v\omega}\sigma_{v\omega}'\Sigma_{v}^{-1}-\beta\sigma_{v\omega}'\Sigma_{v}^{-1}-\Sigma_{v}^{-1}\sigma_{v\omega}\beta'+\beta\beta'\}.$$

Weiter gilt unter Benutzung von Satz 2:

$$\{(I_n,\beta)(I_n,\beta)'\}^{-1} = (I_n + \beta\beta')^{-1}$$

$$= I_n - \frac{\beta\beta'}{1+\beta'\beta}$$

und der Beweis des Satzes 4 gelingt durch Multiplikation von $(I_n,\beta)\Sigma^{-1}(I_n,\beta)'$ in der eben angegebenen Form mit dem Ausdruck auf der rechten Seite der behaupteten Gleichung:

$$\{I_n-\frac{\beta\beta'}{1+\beta'\beta}\}(I_n,\beta)\left(\Sigma - \frac{\Sigma\binom{\beta}{-1}\binom{\beta}{-1}'\Sigma}{\binom{\beta}{-1}'\Sigma\binom{\beta}{-1}}\right)(I_n,\beta)'\{I_n-\frac{\beta\beta'}{1+\beta'\beta}\}=$$

$$= \frac{\beta'\Sigma_{v}\beta\Sigma_{v}-2\sigma_{v\omega}'\beta\Sigma_{v}+\sigma_{\omega\omega}\Sigma_{v}-\Sigma_{v}\beta\beta'\Sigma_{v}+\sigma_{v\omega}\beta'\Sigma_{v}+\Sigma_{v}\beta\sigma_{v\omega}'-\sigma_{v\omega}\sigma_{v\omega}'}{\binom{\beta}{-1}'\Sigma\binom{\beta}{-1}}.$$

Folgerung F3: $\Sigma^{-1}(I_n,\beta)'\{(I_n,\beta)\Sigma^{-1}(I_n,\beta)'\}^{-1}(I_n,\beta)\Sigma^{-1}$

$$= \Sigma^{-1} - \frac{(\overset{\beta}{-1})(\overset{\beta}{-1})'}{(\overset{\beta}{-1})'\Sigma(\overset{\beta}{-1})} \quad .$$

Beweis: Nach Multiplikation der beiden Seiten der Gleichung des Satzes 4 von links mit $(I_n,\beta)'$ und von rechts mit (I_n,β) ergibt sich durch Ausmultiplizieren auf der rechten Seite

$$\Sigma - \frac{\Sigma(\overset{\beta}{-1})(\overset{\beta}{-1})'\Sigma}{(\overset{\beta}{-1})'\Sigma(\overset{\beta}{-1})}$$

und durch Multiplikation mit Σ^{-1} von rechts und links erhalten wir die Folgerung F3.

Folgerung F4: $\text{Spur}(X_2,y^x)\{\Sigma^{-1}-\Sigma^{-1}(I_n,\beta)'[(I_m,\beta)\Sigma^{-1}(I_n,\beta)']^{-1}$

$$(I_n,\beta)\Sigma^{-1}\}(X_2,y^x)' = \frac{(y^x-X_2\beta)'(y^x-X_2\beta)}{(\overset{\beta}{-1})'\Sigma(\overset{\beta}{-1})} \quad .$$

Beweis: Zum Beweis benutze man Folgerung F3 und den

Satz 5: Sind A und B n×m Matrizen, so gilt:

Spur (AB') = Spur (B'A).

Beweis: Man drücke die beiden Spuren durch die Elemente von A und B aus und vergleiche.

Anhang 2: Einige Integralprobleme und damit in Zusammenhang stehende Wahrscheinlichkeitsdichtefunktionen

Erläuterung: Ist $f(x)$ die Dichtefunktion der Zufallsvariablen X so gilt: $\int_R f(x)\,dx = 1$, denn definitionsgemäß ist die Wahrscheinlichkeit des sicheren Ereignisses gleich 1. Hat man nun ein bestimmtes Integral $\int_R g(x)\,dx$ zu berechnen und läßt sich zeigen, daß $g(x) = c\,f(x)$ ist, so gilt trivialerweise $\int_R g(x)\,dx = c$. Im Falle der Anwendung des Bayesschen Theorems läßt dieser einfache Zusammenhang oft die Rückführung komplizierter Integralaufgaben, die bei der Berechnung der Randverteilungen entstehen, auf bereits gelöste Probleme zu. Es folgen einige Beispiele, die in der vorliegenden Arbeit von Bedeutung sind.

1. Normalverteilung

Problem: $$\int = \int_{R^T} e^{-\frac{c_1(y-c_2\xi)'(y-c_2\xi) + c_3(x-c_4\xi)'(x-c_4\xi)}{2\sigma^2}} d\xi$$

wobei ξ, y und x T-Spaltenvektoren und die c_i und σ Skalare sind. Wir bringen die zu integrierende Funktion in die Form: Konstante mal Dichtefunktion einer normalverteilten Zufallsvariablen, indem wir den Zähler des Exponenten nach ξ auflösen und den quadratischen Ausdruck in ξ vervollständigen:

$$c_1(y-\xi c_2)'(y-\xi c_2) + c_3(x-\xi c_4)'(x-\xi c_4)$$

$$= (c_1c_2^2 + c_3c_4^2)(\xi-\hat{\xi})'(\xi-\hat{\xi}) + \frac{c_1c_3(yc_4-xc_2)'(yc_4-xc_2)}{c_1c_2^2 + c_3c_4^2}$$

wobei $$\hat{\xi} = \frac{c_1c_2y + c_3c_4x}{c_1c_2^2 + c_3c_4^2} .$$

Für das zu berechnende Integral gilt also:

$$= (2\pi)^{\frac{T}{2}} \left(\frac{c_1 c_2^2 + c_3 c_4^2}{\sigma^2} \right)^{-\frac{T}{2}} e^{-\frac{c_1 c_3 (yc_4 - xc_2)'(yc_4 - xc_2)}{2\sigma^2 (c_1 c_2^2 + c_3 c_4^2)}}$$

$$\times \int_{\mathbb{R}^T} (2\pi)^{-\frac{T}{2}} \left(\frac{c_1 c_2^2 + c_3 c_4^2}{\sigma^2} \right)^{\frac{T}{2}} e^{-\frac{c_1 c_2^2 + c_3 c_4^2}{2\sigma^2} (\xi - \hat{\xi})'(\xi - \hat{\xi})} d\xi .$$

Die unter dem Integral verbliebene Funktion von ξ ist aber nichts anderes als die Dichtefunktion einer T-dimensionalen Normalverteilung mit dem Mittelwertvektor $\hat{\xi}$ und der Kovarianzmatrix

$$\frac{\sigma^2}{c_1 c_2^2 + c_3 c_4^2} \; I;$$

das Integral selbst ist also gleich 1. Mithin erhalten wir:

$$\int = (2\pi)^{\frac{T}{2}} \left(\frac{c_1 c_2^2 + c_3 c_4^2}{\sigma^2} \right)^{\frac{T}{2}} e^{-\frac{c_1 c_3 (yc_4 - xc_2)'(yc_4 - xc_2)}{2\sigma^2 (c_1 c_2^2 + c_3 c_4^2)}} .$$

1.[x] Multivariate Normalverteilung

Problem: $\int = \int_{\mathbb{R}^m} \sigma^{-\frac{T}{2}} e^{-\frac{(y - X_1\alpha - X_2\beta)'(y - X_1\alpha - X_2\beta)}{2\sigma^2}} d\alpha .$

Kürzen wir ab: $y^{\nabla} = y - X_2\beta$ und vervollständigen den quadratischen Ausdruck in α, so können wir für den Zähler des Exponenten schreiben:

$$\{\alpha - (X_1'X_1)^{-1}X_1'y^{\nabla}\}' X_1'X_1 \{\alpha - (X_1'X_1)^{-1}X_1'y^{\nabla}\} + y^{\nabla\prime} P_1 y^{\nabla}$$

wobei $P_1 = I - X_1(X_1'X_1)^{-1}X_1'$ ist.

Es gilt also:

$$\int = (2\pi)^{\frac{m}{2}} |X_1'X_1|^{-\frac{m}{2}} \sigma^{-\frac{T-m}{2}} e^{-\frac{y^{\nabla\prime} P_1 y^{\nabla}}{2\sigma^2}}$$

$$\int_{R^m} (2\pi)^{-\frac{m}{2}} |X_1'X_1|^{\frac{m}{2}} \sigma^{-\frac{m}{2}} e^{-\frac{\{\alpha-(X_1'X_1)^{-1}X_1'y^{\nabla}\}'X_1'X_1\{\alpha-(X_1'X_1)^{-1}X_1'y^{\nabla}\}}{2\sigma^2}} d\alpha .$$

Der Integrand ist die Dichtefunktion der Normalverteilung mit dem Mittelwert $(X_1' X_1)^{-1}X_1'y^{\nabla}$ und der Kovarianzmatrix $(X_1' X_1)^{-1}\sigma^2$; das Integral ist somit gleich 1. Demnach ist:

$$\int = (2\pi)^{\frac{m}{2}} |X_1' X_1|^{-\frac{m}{2}} \sigma^{-\frac{T-m}{2}} e^{-\frac{(y - X_2\beta)' P_1 (y - X_2\beta)}{2\sigma^2}} .$$

2. χ^2 Verteilung

Problem: $\int_{R_+} \frac{1}{\sigma^{2(\frac{T}{2}+1)}} e^{-\frac{c}{2\sigma^2}} d\sigma^2 = ?$

wobei c und σ Skalare sind. In diesem Fall läßt sich das beschriebene Verfahren anwenden, indem wir substituieren:

$z = \frac{c}{\sigma^2}$ mit $d\sigma^2 = -c z^{-2} dz$ und es gilt:

$$\int_{R_+} \sigma^{-2(\frac{T}{2}+1)} e^{-\frac{c}{2\sigma^2}} d\sigma^2 = c^{-(\frac{T}{2}+1)} \int_0^{\infty} z^{(\frac{T}{2}+1)} e^{-\frac{z}{2}} cz^{-2} dz$$

$$= c^{-\frac{T}{2}} \int_0^{\infty} z^{\frac{T}{2}-1} e^{-\frac{z}{2}} dz.$$

Das Integral über $z^{\frac{T}{2}-1} e^{-\frac{z}{2}}$ von O bis ∞ ist gleich dem reziproken Wert der Normierungskonstante der Dichtefunktion der χ^2 Verteilung mit T Freiheitsgraden. Also gilt:

$$\int_{R_+} \sigma^{-2(\frac{T}{2}+1)} e^{-\frac{c}{2\sigma^2}} d\sigma^2 \propto c^{-\frac{T}{2}}.$$

3. Student-t-Verteilung

Problem: $\int_{\mathbb{R}} \{c_1(x-c_2)^2 + c_3\}^{-\frac{T}{2}} dx = ?$

Zur Lösung transformieren wir hier x in $z^2 = \frac{c_1(x-c_2)^2}{c_3}$,

also $(x-c_2)^2 = \frac{c_3}{c_1} z^2$ und $dx \propto (\frac{c_3}{c_1})^{\frac{1}{2}} dz$ und es gilt:

$$\int_{\mathbb{R}} \{c_1(x-c_2)^2 + c_3\}^{-\frac{T}{2}} dx \propto \frac{c_3^{-\frac{T-1}{2}}}{c_1^{1/2}} \int_{\mathbb{R}} (1+z^2)^{-\frac{T}{2}} dz.$$

$\int_{\mathbb{R}} (1+z^2)^{-T/2} dz$ ist der Kehrwert der Normierungskonstanten der univariaten t-Verteilung. Also gilt:

$$\int_{\mathbb{R}} \{c_1(x-c_2)^2 + c_3\}^{-\frac{T}{2}} dx \propto \frac{c_3^{-\frac{T-1}{2}}}{c_1^{\frac{1}{2}}}.$$

3.* Multivariate t-Verteilung

$$\text{Problem: } f = \int_{\mathbb{R}^m} \left(1 + \frac{\begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}' \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix}\alpha-\hat{\alpha}\\ \beta-\hat{\beta}\end{pmatrix}}{q}\right)^{-\frac{T}{2}} d\alpha.$$

Kürzen wir ab: $\frac{\beta - \hat{\beta}}{q^{1/2}} =: b$ und substituieren α durch

$a := \frac{\alpha - \hat{\alpha}}{q^{1/2}}$ wobei $\alpha = q^{\frac{1}{2}} a + \hat{\alpha}$ und damit die Funktional-

determinante gleich $q^{\frac{m}{2}}$ ist, so gilt:

$$f = q^{\frac{m}{2}} \int_{\mathbb{R}^m} \{1 + (a', b') \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}\}^{-\frac{T}{2}} da.$$

Die quadratische Form in diesem Ausdruck läßt sich wie folgt umformen:

$$Q = (a + (X_1'X_1)^{-1}X_1'X_2 b)' X_1'X_1 (a + (X_1'X_1)^{-1}X_1'X_2 b) + b'(X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2)b .$$

Also gilt mit $P_1 = I - X_1(X_1'X_1)^{-1}X_1'$:

$$\int = q^{\frac{m}{2}}\{1 + b'X_2'P_1X_2b\}^{-\frac{T-m}{2}} \int_{\mathbb{R}^m}\{1 + b'X_2'P_1X_2b\}^{-\frac{m}{2}}$$

$$\left(1 + \frac{\{a + (X_1'X_1)^{-1}X_1'X_2b\}'X_1'X_1\{a + (X_1'X_1)^{-1}X_1'X_2b\}}{1 + b'X_2'P_1X_2b}\right)^{-\frac{T}{2}} da .$$

Der Integrand ist - abgesehen von der Normierungskonstanten, die insbesondere nicht von b abhängt - die bedingte t-Dichte von a gegeben b; das Integral ist also gleich der Inversen dieser Konstanten. Wir erhalten:

$$\int \propto q^{\frac{m}{2}}\left(1 + \frac{(\beta - \hat{\beta})'X_2'P_1X_2(\beta - \hat{\beta})}{q}\right)^{-\frac{T-m}{2}} .$$

Und es gilt weiter:

$$q^{-\frac{T}{2}}\int \propto \{q + (\beta - \hat{\beta})'X_2'P_1X_2(\beta - \hat{\beta})\}^{-\frac{T-m}{2}} .$$

Daß

$$\int_{\mathbb{R}^m}\{1+b'X_2'P_1X_2b\}^{-\frac{m}{2}}\left(1+\frac{\{a+(X_1'X_1)^{-1}X_1'X_2b\}'X_1'X_1\{a+(X_1'X_1)^{-1}X_1'X_2b\}}{1 + b'X_2'P_1X_2b}\right)^{-\frac{T}{2}} da$$

nicht von b abhängt, sieht man leicht mit Hilfe der Substitution:

$$z = \frac{X_1\{a + (X_1'X_1)^{-1}X_1'X_2b\}}{(1 + b'X_2'P_1X_2b)^{1/2}}$$

wobei die Funktionaldeterminante proportional zu

$$(1 + b'X_2'P_1X_2b)^{\frac{m}{2}}$$

ist.

3.^{xx} Matrixvariate t-Verteilung
(Integration von (3.27) über Ξ)

Problem: $\int = \int_{\mathbb{R}^{nT}} |I + (X_2 - \Xi, y - X_1\alpha - \Xi\beta) S^{-1} (X_2 - \Xi, y - X_1\alpha - \Xi\beta)'|^{-\frac{T+k}{2}} d\Xi$.

Wir transformieren: $Z = X_2 - \Xi$ und kürzen ab: $u = y - X_1\alpha - X_2\beta$, so daß:

$$\int = \int_{\mathbb{R}^{nT}} |I + (Z, u + Z\beta) S^{-1} (Z, u + Z\beta)'|^{-\frac{T+k}{2}} dZ.$$

Unterteilen wir S^{-1} wie folgt: $S^{-1} = \begin{pmatrix} S^{v} & s^{v\omega} \\ s^{v\omega}, & s^{\omega\omega} \end{pmatrix}$, wobei

S^{v} eine $n \times n$ Matrix, $s^{v\omega}$ ein n-Spaltenvektor und $s^{\omega\omega}$ ein Skalar ist, so läßt sich der zweite Summand in der Determinante wie folgt weiter umformen:

$$(Z, u+Z\beta) \begin{pmatrix} S^{v} & s^{v\omega} \\ s^{v\omega}, & s^{\omega\omega} \end{pmatrix} (Z, u+Z\beta)' = u(s^{\omega\omega} - q^{12}, (Q^{11})^{-1} q^{12}) u'$$

$$+ \{Z + uq^{12}, (Q^{11})^{-1}\} Q^{11} \{Z + uq^{12}, (Q^{11})^{-1}\}'$$

wobei $Q^{11} = (I_n, \beta) S^{-1} (I_n, \beta)'$ und $q^{12} = (I_n, \beta) \begin{pmatrix} s^{v\omega} \\ s^{\omega\omega} \end{pmatrix}$ ist.

Damit können wir das gesuchte Integral in folgende Form bringen:

$$\int = |I + u(s^{\omega\omega} - q^{12}, (Q^{11})^{-1} q^{12}) u'|^{-\frac{T+k-n}{2}} |Q^{11}|^{-\frac{T}{2}}$$

$$\int_{\mathbb{R}^{nT}} \{|I + u(s^{\omega\omega} - q^{12}, (Q^{11})^{-1} q^{12}) u'|^{\frac{T+k-n}{2}} |Q^{11}|^{\frac{T}{2}}$$

$$\left| I + u\left[s^{\omega\omega} - q^{12}, (Q^{11})^{-1} q^{12}\right] u' + \left[Z + uq^{12}, (Q^{11})^{-1}\right] Q^{11} \left[Z + uq^{12}, (Q^{11})^{-1}\right]' \right|^{-\frac{T+k}{2}} \} dZ.$$

Der Integrand ist - bis auf eine Konstante, die nicht von u oder S abhängt - eine bedingte matrixvariate t-Verteilung für Z bei gegebenem u, das Integral also gerade der reziproke Wert der Normierungskonstanten.

Es gilt weiter:

$|Q^{11}| = |S|^{-1}(\underset{-1}{\beta})' S(\underset{-1}{\beta})$ (vgl. Anhang 1, F2) und

$q^{12\prime}(Q^{11})^{-1}q^{12} = s^{\omega\omega} - \{(\underset{-1}{\beta})' S(\underset{-1}{\beta})\}^{-1}$ (vgl. Anhang 1, F3, wobei nur die letzte Zeile von $S^{-1}(I_n,\beta)'$ und die letzte Spalte von $(I_n,\beta)S^{-1}$ benötigt wird). Damit erhalten wir:

$$s^{\omega\omega} - q^{12\prime}(Q^{11})^{-1}q^{12} = \{(\underset{-1}{\beta})' S(\underset{-1}{\beta})\}^{-1}$$

und folglich gilt:

$$f \propto |S|^{\frac{T}{2}}\{(\underset{-1}{\beta})' S(\underset{-1}{\beta})\}^{-\frac{T}{2}} |I+u\{(\underset{-1}{\beta})' S(\underset{-1}{\beta})\}^{-1}u'|^{-\frac{T+k-n}{2}}$$

oder schließlich:

$$f \propto |S|^{\frac{T}{2}}\{(\underset{-1}{\beta})' S(\underset{-1}{\beta})\}^{\frac{k-n}{2}} \{(\underset{-1}{\beta})' S(\underset{-1}{\beta}) + (y-X_1\alpha-X_2\beta)'(y-X_1\alpha-X_2\beta)\}^{-\frac{T+k-n}{2}}.$$

Literaturverzeichnis

Aigner, D.J.: An Errors in the Variables Model in Which Least Squares is Consistent
International Economic Review 14
(1973) S. 256 und 257

Aigner, D.J.: Errors of Measurements and Least Squares Estimation in a Simple Recursive Model of Dynamic Equilibrium
Econometrica 34
(1966) S. 424-432

Bayes, Rev. T.: An Essay Towards Solving a Problem in the Doctrine of Chances
Philosophical Transactions 53
(1763) S. 376-398 und 401-403 sowie Bd.54
(1764) S. 298-310
wiederabgedruckt in: Biometrica 45
(1958) S. 296-315

Florens, J.P.: Estimation Bayesienne de paramètres des modèles linéaires aléatoires à erreurs sur les variables
Mémoires D.E.S. Faculté de Droit et des Sciences Economiques, Aix-en-Provence
(1969)

Florens,J.P., M.Mouchart und J.F.Richard: Bayesian Inference in Error-in-the-Variables Models
CORE Discussion Paper No. 7221, Louvain
(1972)

Hansen,M.H., W.N.Hurwitz und M.A.Bershad: Measurement Errors in Censuses and Surveys
Bulletin de l'Institut Internationale de Statistique 38,2
(1961) S. 359-374

Hartigan, J.: Invariant Prior Distributions
Annals of Mathematical Statistics 35
(1964) S. 836-845

Jeffreys, H.: Theory of Probability
Oxford
(1966)

Johnston, J.: Econometric Methods
New York - London
(1963)

Kuznets, S.: National Income and Its Computation 1919-1938
National Bureau of Economic Research 40,2
New York
(1954)

Morgenstern, O.: Über die Genauigkeit wirtschaftlicher Beobachtungen
Wien - Würzburg
(1965)

Mouchart, M.: Siehe Florens, J.P.,M.Mouchart und J.F.Richard

Raiffa, H.A. und R.S. Schlaifer: Applied Statistical Decision Theory
Boston: Graduate School of Business Administration, Harvard University
(1961)

Richard, J.F.: Siehe Florens, J.P.,M.Mouchart und J.F.Richard

Rao, C.P.: Linear Statistical Inference and Its Applications
New York - London - Sydney
(1965)

Rényi, A.: Wahrscheinlichkeitsrechnung
mit einem Anhang über Informationstheorie
Berlin
(1962)

Schneeweiß, H.: Ökonometrie
Würzburg - Wien
(1971)

Shannon, C.E.: The Mathematical Theory of Communication
Bell System Technical Journal
(Juli-Oktober 1948) wiederabgedruckt in:
C.E. Shannon und W. Weaver: The Mathematical
Theory of Communication, Urbana (1949)

Studenski, P.: The Income of Nations, Part 2
New York
(1961)

Tiao, G.C.: Siehe A. Zellner und G.C. Tiao

Villegas, C.: On the A Priori Distribution of the
Covariance Matrix
Annals of Mathematical Statistics 40
(1969) S. 1098-1099

Wald, A. The Fitting of Straight Lines if Both
Variables are Subject to Error
Annals of Mathematical Statistics 11
(1940) S. 284-300

Weichselberger, K.: Genauigkeitsansprüche, Fehler und
Kontrollen bei Volkszählungen
Allgemeines Statistisches Archiv 54
(1970) S. 1-28

Zarkovich, S.S.:Quality of Statistical Data
Rom
(1966)

Zellner, A. und G.C. Tiao: Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis
Biometrika 51
(1964) S. 219-230

Zellner, A.: An Introduction to Bayesian Inference in Econometrics
New York
(1971)